Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

COPY 19
210 PAGES

TOP SECRET

SCR 352

25X1

Cy 19 - NPIC/A/TD

C.L. LOGGED 14 MAY 1968

Filed: C&D (IIS) ~~PHS~~ IIS

INTEGRATED INFORMATION SYSTEM (IIS)

PART I - SYSTEM REQUIREMENTS

Vol. 1 - System Operation and Evaluation

NOVEMBER 15, 1967

Prepared for

NATIONAL PHOTOGRAPHIC INTERPRETATION CENTER

First Report of Task Order No. 17

25X1A

25X1

25X1A

DECLASS REVIEW by NIMA/DOD

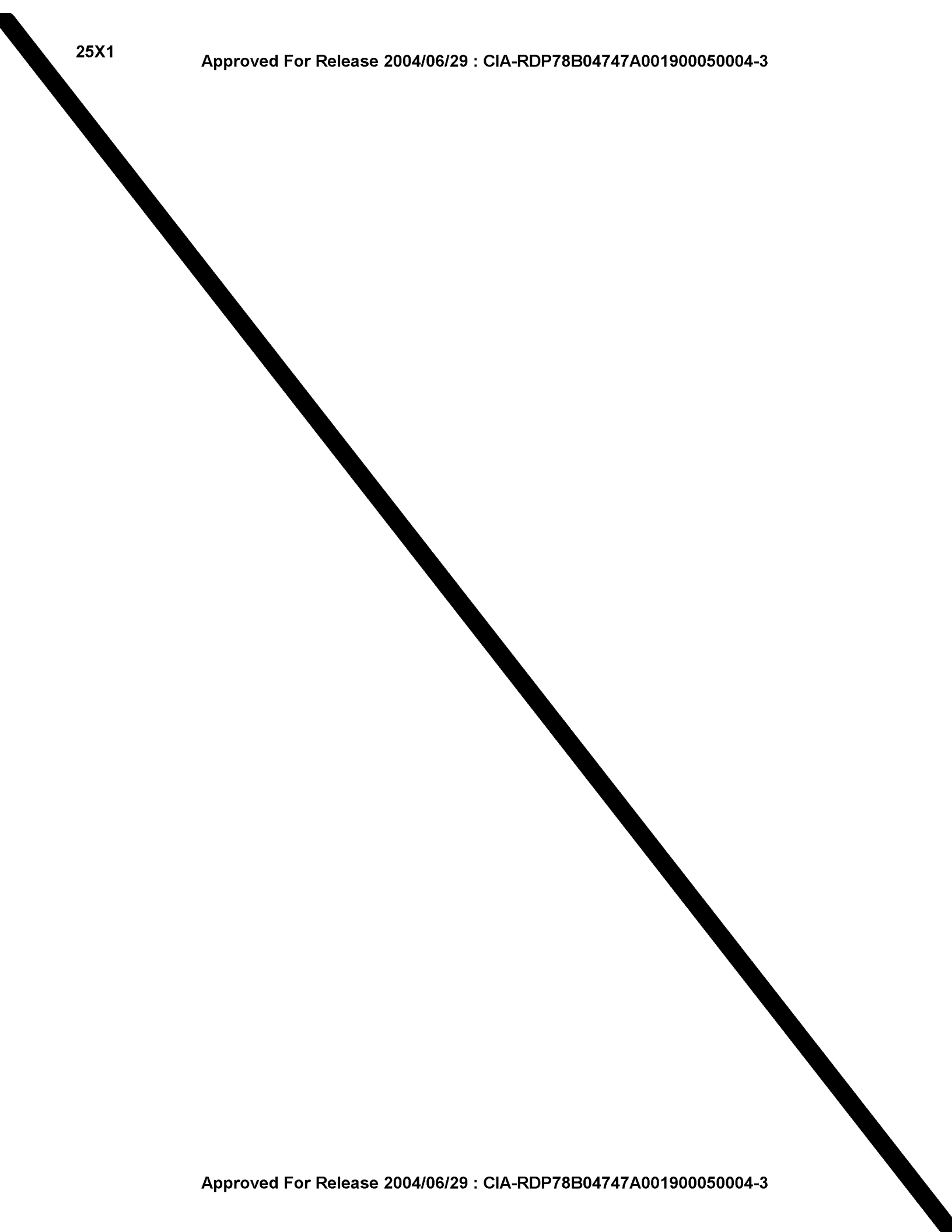
GROUP 1: EXCLUDED FROM AUTOMATIC DOWNGRADING AND DECLASSIFICATION

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

TOP SECRET

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3



Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

SCR 352

INTEGRATED INFORMATION SYSTEM (IIS)

PART I - SYSTEM REQUIREMENTS

VOLUME 1 - SYSTEM OPERATION AND EVALUATION

Prepared for

NATIONAL PHOTOGRAPHIC INTERPRETATION CENTER

First Report of Task Order No. 17

25X1A

November 15, 1967

25X1A

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1 [redacted] TOP SECRET [redacted] 25X1
Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

FOREWORD

This report, in three parts, is submitted under [redacted] 25X1A Task Order No. 17, to provide a detailed description of the Integrated Information System (IIS). Part I, in three volumes, describes the system requirements, the first volume presenting the system operation and evaluation, the second volume presenting the system design, and the third volume presenting the information interface descriptions. Part II, in three volumes, provides the system specifications for software, hardware and files respectively. Part III is a compendium of fifteen individual task memoranda reporting the results of investigative studies performed in support of system design.

25X1
Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3
TOP SECRET [redacted]

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

TABLE OF CONTENTS

Section			Page
1.	SUMMARY		1
	1.1	Background	1
		1.1.1 Basis for the Selected Approach	1
		1.1.2 Brief Design Description	1
	1.2	System Attributes	3
		1.2.1 Increased Capacity	3
		1.2.2 Improved Quality, Timeliness and Reliability	3
		1.2.3 Potential for Future Improvement and Growth	6
		1.2.4 Design Economies	7
		1.2.5 Cost/Effectiveness	8
	1.3	IIS Description	8
		1.3.1 Requirements	8
		1.3.2 System Configuration	9
		1.3.3 System Operation	21
	1.4	Report Organization	25
2.	ROLE OF THE CENTER WITHIN THE COMMUNITY		27
	2.1	Immediate Readout	27
	2.2	Detailed Reporting	28
	2.3	National Data Base Responsibilities	28
	2.4	Summary of Major Changes	29
3.	INPUT-OUTPUT LOADING FACTORS		31
	3.1	Introduction	31
		3.1.1 Purpose	31
		3.1.2 Data Sources	31
		3.1.3 Organization	31
	3.2	Parameters Related to Immediate Reporting	32
		3.2.1 Missions/Year	32
		3.2.2 Feet of Film/Mission	34
		3.2.3 Targets Covered/Mission	34
	3.3	Parameters Related to Basic Intelligence Support Activities	35
		3.3.1 Purpose	35
		3.3.2 Sources of Data	35
	3.4	Parameters Related to the Data Base Size	35
		3.4.1 Purpose	36
		3.4.2 Sources of Data	36
	3.5	Data Base Activity	37
		3.5.1 Purpose	37
		3.5.2 Sources of Data	37

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

TABLE OF CONTENTS (cont'd)

Section		Page
4.	OVERALL SYSTEM OPERATION	39
4.1	System Configuration	39
4.1.1	Computer Complex	39
4.1.2	Consoles	40
4.1.3	Software System	44
4.2	Data Base Organization	47
4.2.1	Installations Data File (IDF)	48
4.2.2	Moving Target File (MTF)	50
4.2.3	Objects Data File (ODF)	50
4.2.4	Exploitation Products Data File (EPDF)	51
4.2.5	Mensuration Parameters File (MPF)	52
4.2.6	File Indices	52
4.3	Immediate Readout Activities	53
4.3.1	Mission Readiness	53
4.3.2	Readout Activity	56
4.3.3	Output Cycle	60
4.4	Detailed Reporting	62
4.4.1	Preparation	62
4.4.2	Readout	63
4.4.3	Report Preparation	65
4.5	National Data Base Activities	65
4.5.1	Data Base Inputs	65
4.5.2	Query Capability	68
4.5.3	COINS Interface	75
4.5.4	Data Base Maintenance	76
5.	DESIGN STUDIES, ANALYSIS AND ALTERNATIVES	79
5.1	Data Base Considerations	79
5.1.1	Size Estimates	79
5.1.2	Utilization	86
5.1.3	Storage Media	90
5.2	PI Presentation Sheet Content	97
5.3	First and Second Phase Studies	103
5.3.1	Cost vs. Time	103
5.3.2	Console Reliability	113
5.4	System Capacity and Queues	113
5.5	Input/Output Console Study	119
5.5.1	Determination of Console Types and Numbers	119
5.5.2	Analysis of Console Utilization	121
5.5.3	Conclusions	124
5.6	Language Considerations and Alternatives	124
5.6.1	Query Language Considerations and Alternatives	124
5.6.2	Program Language Alternatives in IIS Implementation	131

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

TABLE OF CONTENTS (cont'd)

Section		Page
5.7	Cost Benefits to Pre-Readout Activities	135
5.7.1	Basic Operations Under Analysis.	136
5.7.2	Interrelationships of Operations	136
5.7.3	Products of Operation.	136
5.7.4	Product Cost Analysis.	137
5.7.5	Conclusions.	138
5.8	Data Base Reliability	138
5.8.1	Detection Methods.	138
5.8.2	Requirements for Error Detection	139
5.8.3	Independent File	141
5.8.4	Analysis of Errors	143
5.8.5	Storage Hardware Maintenance	144
5.9	UNIVAC 490/494 Software Survey.	144
5.9.1	UNIVAC 490 Series Installations.	144
5.9.2	Installations with Similar Applications.	145
5.9.3	Survey Results	145
6.	COST/EFFECTIVENESS EVALUATION.	149
6.1	Measures of Cost/Effectiveness.	149
6.1.1	Cost Factors	149
6.1.2	Effectiveness Factors.	150
6.1.3	Composite Measure.	153
6.2	Extrapolation of Current System	155
6.2.1	Personnel.	155
6.2.2	Equipment Costs.	156
6.2.3	Performance of the Extrapolated System	160
6.3	IIS Cost and Performance.	162
6.3.1	IIS Personnel.	162
6.3.2	IIS Equipment Costs.	162
6.3.3	Performance of the IIS System.	162
6.4	Cost Effectiveness Ratios	166
APPENDIX A -	Analysis of Routine Dissemination/Request	A-1
APPENDIX B -	Analysis of Data Base Utilization	B-1
APPENDIX C -	Tape Storage System	C-1
APPENDIX D -	Drum Storage System	D-1
APPENDIX E -	System Model.	E-1
APPENDIX F -	Reliability vs. Lost Time	F-1
APPENDIX G -	Queues and Delays	G-1

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

LIST OF TABLES

Table		Page
I	Summary Manpower Estimates by Major Category	4
II	Mission Oriented Activities (Immediate Reporting)	10
III	Basic Intelligence Support Activities (Detailed Reporting)	11
IV	Data Base Size Estimates	12
V	Data Base Activity Estimates	13
VI	Input/Output Equipment	14
VII	IDF Size Estimate	80
VIII	Use of IIS Data Base	87
IX	Data Utilization - First Phase Readout	98
X	Data Utilization - Second Phase Readout	99
XI	PI Presentation Characteristics	102
XII	Ranges of Parameters	104
XIII	Average Waiting Time in Queue	117
XIV	IIS Console Characteristics	120
XV	Hypothetical Manpower Scheduling (Constant Support)	122
XVI	Hypothetical Manpower Scheduling (Constant PI's)	123
XVII	List of Representative Activities	157
XVIII	Manpower Extrapolation Factors	159
XIX	Manpower Summary and Extrapolation	159
XX	Summary of Extrapolated System Equipment Costs	161
XXI	Selected Manhour Estimates by Major Activity	163
XXII	Summary of IIS Equipment Costs	164

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

LIST OF FIGURES

Figure		Page
1	Hardware Configuration in the 1971 NPIC Computer Environment	15
2	User Interface with Data Base	18
3	Index Structure	20
4	Role of NPIC in the Community	22
5	Initial IIS Hardware Configuration (1968)	41
6	Mission Readiness Activities.	54
7	Immediate Readout	57
8	Development Objectives - Presentation and Worksheet	58
9	Output Cycle.	61
10	Detailed Reporting.	64
11	Query Language Terms.	70
12	Automatic Sources of Data Base Updates.	77
13	Manual Sources of Data Base Updates	78
14	Data Base Size Estimates.	85
15	Data Base Size Breakdown.	85
16	Use of IIS Data Base.	89
17	Characteristics of Tape Storage	93
18	Characteristics of Drum Storage	94
19	Storage Comparison.	96
20	Effect of Varying Presentation Content.	100
21	Manpower Estimates for Phase I Reporting.	106
22	Manpower Estimates for Phase II Reporting	108
23	Estimates for Immediate Reporting Costs (Constant PI's)	110
24	Estimates for Immediate Reporting Costs (Constant Support).	111
25	Estimates for Immediate Reporting Costs (Constant PI's with Consoles for Support Only).	112
26	Product Time Lost/Year.	114
27	Expected Queues for Storage Access.	116
28	Expected Queues Under Priority Conditions	118
29	IIS Query Capabilities.	130
30	Effectiveness of Sampling	142
31	490 Series User Survey.	146
32	Effect of Varying Weights of Time Factors	167
33	Effect of Varying Cost Weighting Factors.	168

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

1. SUMMARY

1.1 Background

1.1.1 Requirement for an Integrated Information System (IIS)

NPIC has the problem of developing the maximum amount of useful information from the imagery of overhead photography. This demands a continuing improvement in techniques and practices employed as well as in efficiency. At the heart of the NPIC operation is the work of the photo-interpreter himself. The key to improving his product lies in the completeness and timeliness of the support available to him. In the spring of 1966, NPIC established a program to explore the use of computer techniques both to improve the present product and to provide a means for dealing with the future projected load in a cost/effective manner.

A survey of NPIC in the light of current computer technology indicated that it was possible to store most of the information essential for NPIC operations in the memory of a computer, to which Center personnel could have direct (on-line) and rapid access. The obvious advantages of such an Integrated Information System (IIS) for the Center are immediacy and convenience of data support to the photo-interpreter and immediacy of data update. For the intelligence community, IIS means more rapid dissemination of reports and, at the same time, relief from complete dependence upon reports by the provision of channels for direct inquiry. To both the Center and the community, the "on-line" aspect of an IIS offers a powerful means of concentration on areas of critical interest in times of emergency. Thus, only through an integrated information system may it be possible for NPIC to effectively meet the obligation assigned to it by the National Tasking Plan of maintaining a national data base.

1.1.2 Brief Design Description

It was decided after a general approach was selected, to bring the system into being in four consecutive work phases: conceptual design, detailed design, implementation, and installation and test, with operation planned for late 1968. The conceptual design was completed in 1966. The subsequent detailed design phase was begun in December 1966 and completed in October 1967. Its results are described in the body of this report.

Briefly the design calls for a data base consisting of a mensuration parameters file, an installations (target-brief) file, an objects file, and the all-source index to be stored in high-speed, random access magnetic drum memories. The drum memories are controlled by a real-time computer. Center personnel, depending upon their function, interact with the computer/storage complex through a variety of terminal units varying from simple

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

keyboards to enter data to cathode ray tubes for display. High speed printers are provided to produce outputs not involving interaction with the computer in real-time. An interface with other computer systems is included in the design, and communications line terminals are provided to accommodate this interface and to permit other remote outputs.

The functions performed by the IIS contribute to Center operations preparatory to readout, during readout and following readout. Prior to readout the system predicts targets covered, provides data to aid in target assignment and scheduling, issues instructions for film control, and provides basic support data. During readout the system allows the photo-interpreter to query the data base for additional information. It also allows for on-line mensuration. Following readout the system facilitates entry, verification, edit, and approval of results. The approved results are used to update the data base automatically and generate NPIC products. Distribution information is also automatically provided.

The philosophy adhered to in selecting functions to be automated was to stay well within the state-of-the-art in both software and hardware. By definition the IIS was limited in scope to "a computer system for handling alphanumeric substantive data and indices to substantive data." No attempt was made to perform direct digital operations on imagery nor to automate film handling. The computers already in place at NPIC are completely adequate as a base upon which to build the system. A suitable software system exists in the form of the RUM Executive System. None of the terminal devices called for require costly development.

Within the design constraints, all functions proposed are readily achievable. Without an IIS there would be a requirement for an 80% increase of critically short photo-interpreters by 1971, and even with this added manpower, it would be difficult to realize the new dimension of intelligence that should be expected to result from more comprehensive and frequent coverage. Automation of the Center's information processes through the IIS will keep the augmentation of the Center's staff to a minimum and will permit more timely and sophisticated interpretation of intelligence data. The great significance of the IIS is that it represents a concrete first step in the dynamic use of computers for imagery exploitation. As such, it may, as in the case of weapons design and cryptography, be a first step toward a new and higher order of operation that would be impossible except through automation.

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

1.2 System Attributes

This section reviews the attributes of the IIS principally from the point of view of their economic and performance advantages. The IIS is discussed in terms of:

a) The increased capacity it provides for imagery exploitation
b) The improved quality, timeliness, and reliability it permits
c) The potential it provides for future improvements, growth, and greater depth of interpretation
d) Its design economies
e) Its cost/effectiveness

1.2.1 Increased Capacity

The IIS increases the capacity of the Center for imagery exploitation by improving the utilization of PI's and by simplifying or eliminating support functions. By 1971 NPIC, whose capacity and supply of PI's is currently strained, will be faced with an 80% increase in its workload. With the support of the IIS, it is estimated that this increased load will require that the number of PI's be increased by only 50% and that no increase will be required in the number of support personnel. If current procedures are retained, it is estimated that the number of PI's will have to be increased in direct proportion to the workload and that approximately 60 additional support personnel will be required. Table I summarizes manpower requirements for the current and proposed (IIS) systems.

The capacity of the IIS is not limited by the workload projected for the 1971 time frame. Because of the modularity of the system at least a doubling of the 1971 workload could be accommodated, without a basic system redesign, through the addition of memory and terminal equipment.

1.2.2 Improved Quality, Timeliness and Reliability

The intellectual process involved in photo interpretation is advanced in thoroughness and reliability directly by the IIS. The PI's whole attitude and efficiency are influenced by the readiness and completeness of the response which he gets to his need for support. He is also influenced by an improvement in the attractiveness of his environment.

On the simplest level, the Presentation Sheets to be furnished him (as a modification of the current Target Brief File Blip Sheet) have been purged of redundant and irrelevant information and are conveniently organized.

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

TABLE I

SUMMARY MANPOWER ESTIMATES BY MAJOR CATEGORY

Catagory	FY 1967 *	Extrapolated System *	IIS *
Photo Interpreter	153	280	224
Support	76	136	75
IPD	55	80	43 **
Total	284	496	342

* Figures are in people.

** IPD people are 5-year average of personnel, directly supporting the IIS, i.e. 28 programers + 15 computer operators.

TOP SECRET

TOP SECRET

25X1

25X1

SCR 352

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

When he needs additional information, his search of the data base is facilitated by a query capability with a response time in seconds allowing him to carry on a dialog with the computer without interruption. Since information is available to the PI within a given time period, the probability of his overlooking pertinent information is reduced. The worksheet used for reporting readout is designed to minimize the amount of textual composition by the PI through the use of checkmark and line correction techniques and to facilitate entry into the data base without further interpretation.

Time to update the data base following immediate readout has been decreased by a factor of ten, thereby improving the currency of information. This is accomplished by entering the readout data for a given target immediately upon approval, rather than batch processing some two weeks later as is done currently.

Automated techniques, in place of current manual techniques, permit the mission readiness activities of target prediction, preparation of PI support material, assignment of targets to PI teams, and determination of film distribution to be accomplished in hours rather than days. This capability, while not required under normal operating conditions, will be necessary during crisis situations and in the advent of direct readout satellites or on-board processing on high-speed aircraft missions.

The use of on-line consoles and random access storage in the production cycle provides two major advantages. First, since the information remains in machine readable form throughout the cycle, there is less paper passing, less elapsed time between operations, less reentry of data, and less chance of error. Second, since the output products are assembled directly from computer-stored data, there is a great deal of flexibility for changing content and organization of the products.

The reliability of the data is enhanced by the use of computer programs which keep track of data transactions to assist in determing the location and extent of errors and to expedite their correction.

The IIS interface with COINS will provide rapid exchange of information between the Center and other elements of the community. The structured nature of the IIS data base and the provision for selective retrieval of data will allow agencies outside the Center to correlate imagery-derived information with other types of data, using computer methods. The value of correlation has been demonstrated, but it has not been used extensively because of the large amount of manpower required using current methods.

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

1.2.3 Potential for Future Improvement and Growth

Change in the imagery exploitation process is inevitable. Accordingly, the ability to accommodate change has been a paramount consideration in the design of the IIS. A number of the ways in which the IIS can adapt to (or stimulate) change are described below.

a) IIS Internal Operations - Since the IIS computer is an integral part of the path of information flow through the Center, a computer program is provided to generate statistics on this data flow. In this way impediments to the flow of information and areas of potential improvement can be identified.

b) Materials Handling - In recognition of a need to provide improved physical handling of materials such as film, reports, drawings, and other collateral items, the IIS query capabilities will provide the means of identifying the pertinent materials through the various indices and references to materials that are carried in the data base.

c) New Exploitation Techniques - Automatic or semiautomatic aids to the PI such as target recognition and change detection, when operationally useful, will likely require the availability of quantitative measures of target characteristics (i.e., counts, sizes, shapes, orientations). These techniques can be accommodated in the IIS by incorporating the quantitative measures in the data base and making them available on request, much as other kinds of data are made available to the PI during readout.

d) New Collection Systems - The advent of direct readout satellites or on-board processing on high speed aircraft missions will require rapid preparation for readout. The automated IIS procedures for target prediction, preparation of PI support material, assignment of targets to PI teams, and determination of film distribution will be essential for reducing the time required for achieving a state of readiness.

e) Data Base Expansion or Reorganization - If the content or size of the data base should increase for any reason beyond that projected for the 1971 time period, the structure of the data base and modular characteristics of the storage media allow for the expansion or addition of new files, sectors, and fields and the modification or creation of new indices.

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

f) Product Change - The content, organization, and format of reports, cables, and other products provided by the IIS can be readily changed when necessary through use of the IIS Presentation computer program.

1.2.4 Design Economies

A ground rule observed throughout the design of the IIS was to incorporate available software and hardware which were also consistent with meeting performance objectives. Consequently, software development was kept to a minimum and it was possible to avoid new hardware development entirely.

1.2.4.1 Software

The RUM Executive System developed by NSA and currently used by the Center was adopted for control of the IIS. Choice of the RUM Executive System permits the Center to continue to use a large number of non-IIS applications programs without reprogramming.

25X1A The performance and costs to develop and maintain the TIPS (NSA), CAPRI (CIA), and PIRL [redacted] query languages were analyzed. [redacted] 25X1A language was selected over TIPS because of superior performance (but equal cost), and over CAPRI because of the lower cost (but equal performance).

On-line and batch mensuration programs in use or under development by the Center will be utilized in the IIS.

UNIVAC 494 Assembler and FORTRAN and COBOL compilers developed by NSA will be used for computer program preparation in preference to developing a new IIS System Language compiler.

In addition to consideration given to the utility of the NSA computer programs mentioned above, a survey of UNIVAC 490/494 installations was conducted to determine the existence, suitability, and availability of software to serve the needs of the IIS. The level of effort expended for the survey was not large, but provided the desired assurance that potential savings in IIS software implementation were not overlooked.

1.2.4.2 Hardware

Dual UNIVAC 494's, with their associated peripherals as presently in use or planned for the Center were selected as the IIS computer complex. This equipment embodies the required performance characteristics, and its use eliminates the need to convert existing Center software, retrain personnel, or operate a multi-computer complex during transition from the current system to the IIS. The associated savings achieved outweigh any

25X1 [redacted] TOP SECRET Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3 [redacted] 25X1

possible economies gained from the use of other hardware over the life of the IIS.

No new acquisitions of plotting equipment are planned. The variety of computer or magnetic tape driven plotters already owned by the Center will meet the IIS requirements for graphic outputs.

In the specification of new equipment, commonality to reduce costs and maintenance problems and the use of equipment with multiple capabilities to improve utilization has been a principal design consideration. No costly equipment R&D is required since all units specified are commercially available.

1.2.5 Cost/Effectiveness

A comparison of the proposed mode of Center operation with an extrapolation to 1971 of its mode of operation under the current system shows that the IIS will result in a 19 to 1 reduction in costs per unit of effectiveness. This factor derives in large part from several major improvements, namely the ability to:

a) Produce the OAK in two days
b) Update the data base in five minutes
c) Respond to a query in six seconds

The times listed above contrast to three days, four hours, and ten minutes, 25X1A respectively, for the extrapolated current system.

[redacted]

1.3 IIS Description

The IIS can be described conveniently in terms of its requirements, configuration, and operation.

1.3.1 Requirements

The requirements to be fulfilled by the IIS i.e., the necessary system capabilities and input/output loading data, are the basis for the system design. It should be noted that the input/output loading data do not represent a capacity limit for the system, but serve as 1) a design point at which the system will probably operate during the 1969-1974 time period,

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3 [redacted]
TOP SECRET

and 2) as a basis for performance comparison of the IIS with an extrapolation of the current system. Wherever possible, data were drawn from authoritative sources such as the National Tasking Plan and annexes to the JIIRG Report, but in all cases the values were submitted to and concurred in by the Center.

Capabilities - The IIS must:

a) Provide capacity for the imagery exploitation workload projected for the 1969-1974 time period.
b) Improve the quality and timeliness of imagery exploitation.
c) Provide growth potential and flexibility to accommodate change.
d) Be compatible with the community and national intelligence program.
e) Provide cost/effectiveness exceeding the extrapolated current system.

Input/Output Loads

The quantitative data on input/output loads are shown in Tables II, III, IV and V. These data have been organized in relation to their effect on Immediate and Detailed Reporting, the size of the data base, and loads imposed on the system by data base activity.

1.3.2 System Configuration

The system configuration includes hardware, software, and the data base (files). The configuration as herein described is for the fully implemented system.

1.3.2.1 Hardware

The computer complex will consist of dual UNIVAC 494's with their associated peripherals and will develop in an evolutionary sense from the present NPIC configuration. The remote man/machine interface units and their use, quantity, and distribution throughout NPIC are shown in Table VI. The overall hardware configuration is as shown in Figure 1.

This configuration provides for mass storage of a major portion of the data based on Fastrand II drums, a principal factor in enabling rapid response to data base queries. Various kinds of consoles are provided for entering and checking the data and for querying the data base. Mensuration site plots, engineering drawings, and mission coverage plots will be produced on plotters already owned by NPIC. All input/output equipment has upper and lower case capability. Multiple units of key hardware components

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

TABLE II

MISSION ORIENTED ACTIVITIES (NORMAL/MAXIMUM)

(IMMEDIATE REPORTING)

PARAMETER \ TYPE OF MISSION	ORBITAL SEARCH	ORBITAL SPOTTING	AIRCRAFT PHASE I & II	AIRCRAFT PHASE II ONLY
1. NUMBER OF MISSIONS/YEAR	15* / 20**	8 / 12	100 / 150	500 / 1000
2. FEET OF FILM/MISSION	30,000	3,000	5000 / 7000	5000 / 7000
3. TARGETS COVERED/MISSION				
a) High Priority COMIREX	250 / 400	450 / 600	17 / 18	
b) Low Priority COMIREX	750 / 1100	1350 / 1600	51 / 54	
c) Total	4000 / 6000	2700 / 3300	198 / 210	198 / 210

* Normal
** Maximum

TOP SECRET

TOP SECRET

25X1

25X1

SCR 352

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

TABLE III

BASIC INTELLIGENCE SUPPORT ACTIVITIES
(DETAILED REPORTING)

PARAMETER	VALUE (PER ANNUM) NORMAL	MAXIMUM
A. TOTAL NUMBER OF INSTALLATIONS TO BE COVERED	10,100	12,000
B. RATE OF PREPARATION OF INITIAL REPORTS	2,100	3,000
C. RATE OF PREPARATION OF NPIC INITIAL REPORTS	300	700
D. TOTAL UPDATE RATE	5,000	6,300
E. NPIC UPDATE RATE	1,400	1,900
F. PUBLICATION RATE OF NON-NPIC INITIAL REPORTS	1,800	2,300
G. PUBLICATION RATE OF NON-NPIC UPDATE REPORTS	3,600	4,400

TOP SECRET

25X1

25X1

SCR 352

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

TABLE IV

DATA BASE SIZE ESTIMATES

PARAMETER	VALUE
A. INSTALLATIONS DATA FILE	
COMIREX Targets (Active)	6,000
NIETB Total	36,000
NPIC Total	42,000
B. OBJECTS DATA FILE	
Total Number of Objects	1,000
No. on Which "Where Seen" Information is Kept	All
Average Number of Locations	250
C. EXPLOITATION PRODUCTS DATA FILE	
Number of Reports Currently Included	80,000
Growth Rate	11,000/Year

TOP SECRET

25X1

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

TABLE V

DATA BASE ACTIVITY ESTIMATES

	ACTIVITY	RATE NORMAL	RATE MAXIMUM
A.	COINS		
	Number of Queries/Day	1,000	2,000
	Average No. of Characters/Response	2,000	50,000
B.	NUMBER OF PHASE I AND PHASE II REPORTS RECEIVED FROM OTHER AGENCIES/DAY	60	120
C.	NUMBER OF CHANGES IN REQUIREMENTS IN THE IDF AND ODF/DAY	25	50
D.	NUMBER OF NEW COLLATERAL DATA ITEMS FOR THE IDF AND ODF/DAY	25	50
E.	RATE OF NEW INSTALLATION ACQUISITION/DAY	16	50

TOP SECRET

TOP SECRET

25X1

25X1

SCR 352

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

TABLE VI

INPUT/OUTPUT EQUIPMENT

UNIT	USE	LOCATION	QUANTITY
Edit Console U/L Cs	Verify	PAG	7
Edit Console U/L Cs	Edit	PAG	3
Edit Console U/L Cs	Approve	PAG	2
Query Console	SUSIE	PAG	5
TTY-Printer/Kbd.	PI	PAG	10
TTY Printer	PI	PAG	10
TTY Kbd.	Data Entry	PAG	4
Plotter 635	Graphics	PAG	1
Edit Console U/L Cs	Data Entry	CSD	3
Edit Console U/L Cs	Verify	CSD	2
TTY Printer	*	CSD	2
Query Console	SUSIE	CSD	2
Edit Console U/L Cs	Edit	PD	6
TTY Kbd.	Data Entry	PD	3
Auto. TTY^{W}/PUNCH	Cables	OS	1
Printer 600 LPM	Pg.Printing	Computer Room	1
Card Reader	MPF	Computer Room	1
Paper Tape Punch	Reports	Computer Room	1
TTY Printer/Kbd.	Film Records	PSD	2
TTY Printer	*	IAS	1
Plotter 635	Graphics	IAS	2
TTY Printer/Kbd. Model H	Mens.	TID	6
Plotter Model LTE	Graphics	TID	1 each
TTY Printer	*	DIAAP	1
Plotter Model LTE	Graphics	DIAAP	1
Plotter 635	Graphics	IPD	2
Plotter E-80	Graphics	IPD	1
TTY-Printer/Kbd.	Programming	IPD	1

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

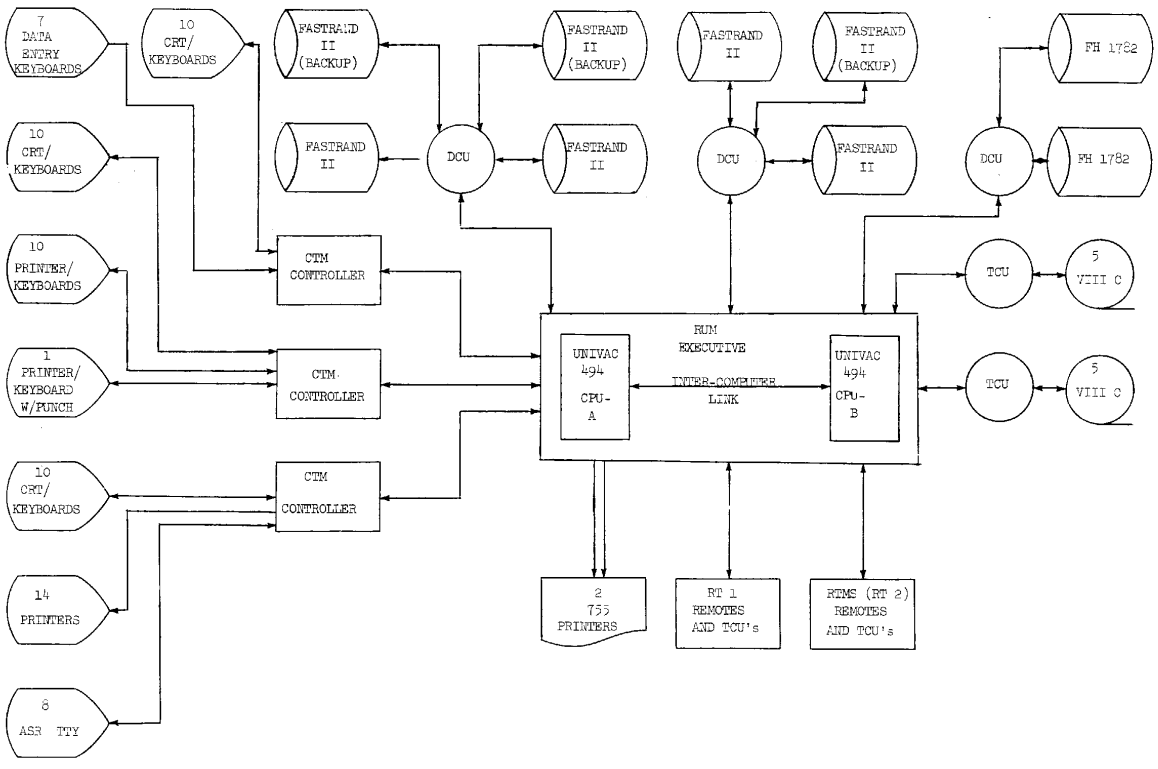


Figure 1 Hardware Configuration in the 1971 NPIC Computer Environment

TOP SECRET

TOP SECRET

25X1

25X1

SCR 352

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3
25X1

are provided for backup purposes, however interchangeability of console functions holds redundancy to a minimum.

1.3.2.2 Software

The software complement which drives the hardware, and in turn provides the capabilities of the IIS, consists of computer programs to support 16 functional areas. These areas are as follows:

a) File Update and Maintenance - Creation of sections of, addition of data to, and deletion of data from the operational data base.

b) Data Entry - Acceptance of input data from various equipment, (keyboard, card reader, etc.), error checking, conversion and formatting of the data for further processing.

c) Verification, Editing and Approval - Support the VEA (Verify, Edit and Approval) function, including the retrieval, display and modification of data in response to commands from consoles manned by verifiers, editors, and the approving authority.

d) On-Line Mensuration - Support to real-time mensuration for PAG, TID, or resident detachments.

e) Query Handling - Decoding of a user-oriented Query Language and retrieval of desired sets of information from the operational data base.

f) Cables and Real-Time Reports - Formatting and preparation for transmission of either computer-to-computer information or outgoing cables.

g) Report Generators - Generation of hard copy reports automatically from information stored in the data base.

h) Executive - Control of the entire IIS software system, including all I/O handling, program scheduling, security, maintenance, etc.

i) Update Mensuration Parameter File - Generation of the necessary parameters, on a mission basis, to support both on- and off-line mensuration.

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

j) Batch Mensuration Support - Similar to on-line mensuration (d) except that support is provided by the IIS in an off-line sense generally using punched cards as input.

k) Target Prediction - Prediction of targets covered by a given mission prior to film arrival.

l) Assignment - Assignment of film and necessary support materials to various PI teams for Phase I and Phase II readout.

m) Program Languages - Provision of assembler and compilers to be used to generate all of the information handling software for the IIS.

n) File Maintenance Language - Processing of an on-line input language similar to the Query Language (e). The language provides the user interface with the maintenance procedures themselves (a) e.g. CSD corrections to Header Information or the addition of collateral material to the Installations Data File.

o) Presentation Language - Processing of an on-line language which permits the user to change or create report formats.

p) Miscellaneous Software - Provision of a collection of programs to support the photo reference file, maintenance and system operations.

Control of the IIS will be accomplished by an IIS Real-Time program and a File Control Program operating under the RUM Executive Control system currently in use at the Center.

Logical packaging of the programs to support the functional areas and their employment when a user interfaces with the data base is shown in Figure 2.

1.3.2.3 Data Base

The data base consists of information on installations, objects, exploitation products, and mensuration. It serves to support the operations of the Center and as a national repository for imagery derived information in accordance with the National Tasking Plan.

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

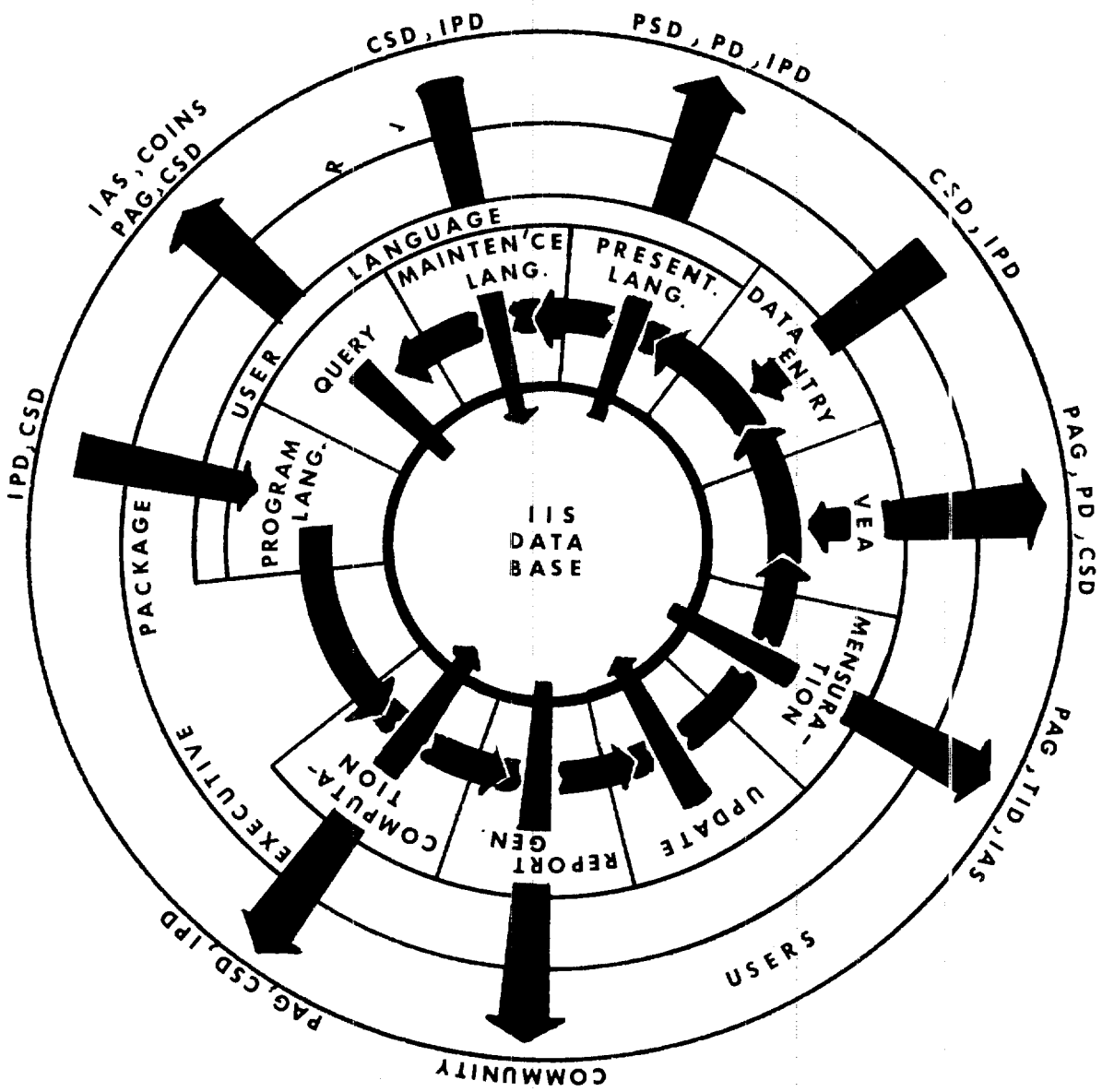


Figure 2 User Interface With Data Base

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

The physical storage of data among the various storage media places the most used data in the most accessible storage media and the less used data in progressively less accessible storage. Assignments can be conveniently changed in accordance with experience gained during system operation.

The data base is logically divided into files, each file into unit records, and each unit record into sectors. The five major files are as follows:

Installations Data File (IDF)

Analogous to the current Target Brief File. It will eventually contain one unit record for each of 42,000 installations.

Moving Target File (MTF)

Consists of a unit record for each unique mobile object of primary interest to the intelligence community and provides the characteristics and historic record of movement.

Objects Data File (ODF)

Consists of a unit record for each object. An all-source collection of information is provided on military weapons, logistic equipment, significant R&D, communications, industrial and transportation components and equipment.

Exploitation Products Data File

Analogous to the All-Source Index currently maintained by the Center.

Mensuration Parameters File (MPF)

Consists of a unit record for each mission and provides all the parameters necessary to derive measurements from imagery.

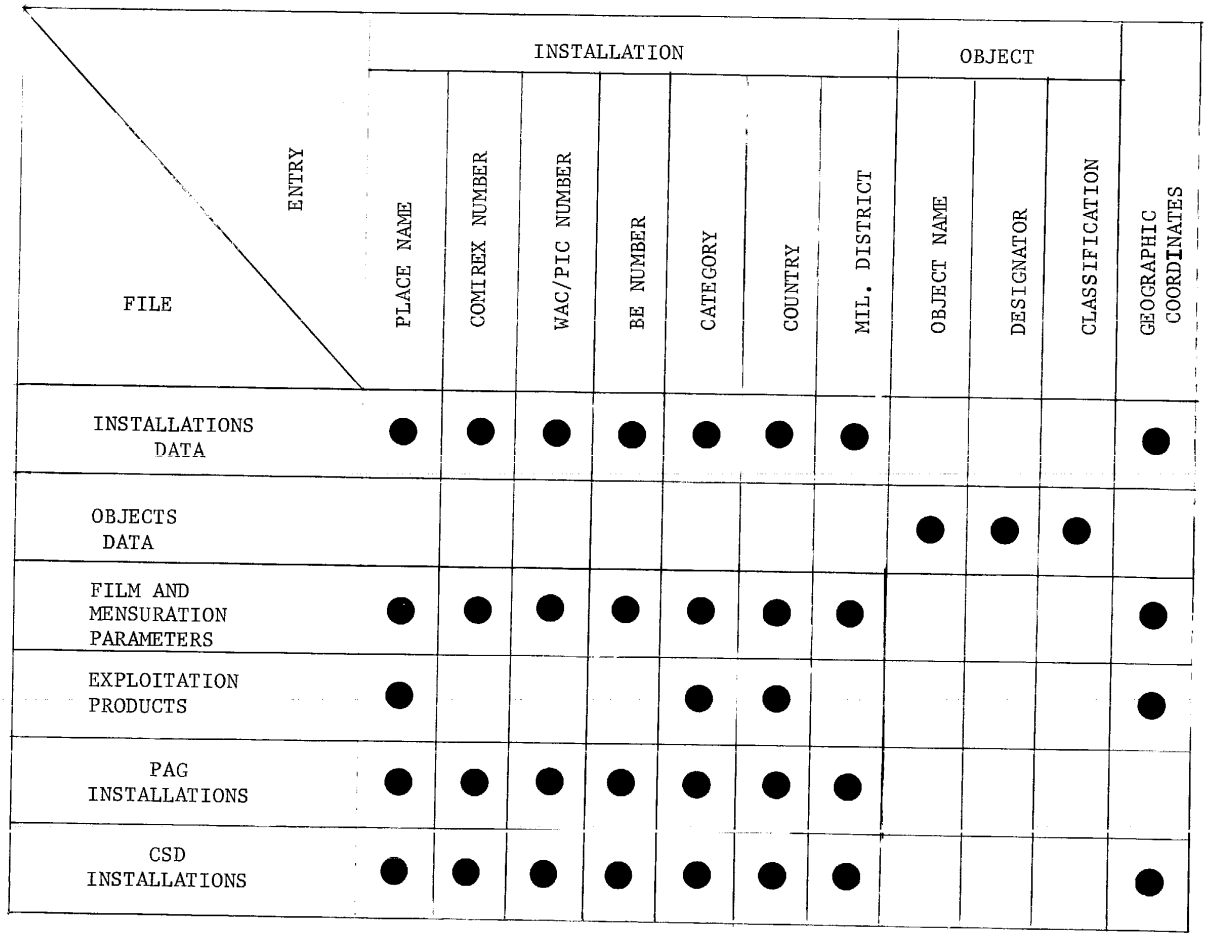
Rapid search of the files can be achieved through the use of any of the commonly used access terms. This is facilitated by the availability of indices to relate the access terms to the location of the index records of the corresponding unit records. Indices can be created on the basis of any of the terms in the data base. The indices planned for initial implementation along with the files to which they provide access are shown in Figure 3.

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1



FILE \ ENTRY	INSTALLATION							OBJECT			
	PLACE NAME	COMIREX NUMBER	WAC/PIC NUMBER	BE NUMBER	CATEGORY	COUNTRY	MIL. DISTRICT	OBJECT NAME	DESIGNATOR	CLASSIFICATION	GEOGRAPHIC COORDINATES
INSTALLATIONS DATA	●	●	●	●	●	●	●				●
OBJECTS DATA								●	●	●	
FILM AND MENSURATION PARAMETERS	●	●	●	●	●	●	●				●
EXPLOITATION PRODUCTS	●				●	●					●
PAG INSTALLATIONS	●	●	●	●	●	●	●				
CSD INSTALLATIONS	●	●	●	●	●	●	●				●

Figure 3 Index Structure

TOP SECRET

25X1

25X1

TOP SECRET

SCR 352

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

1.3.3 System Operation

The IIS is designed to enhance the capability of the Center to carry out its responsibilities in the intelligence community for immediate reporting of certain photographic missions, detailed reporting on certain categories of installations, and maintenance of the National Data Base of Imagery-Derived Information concerned with installations, objects, exploitation products, and data required for mensuration. These responsibilities and their relationship to the community are as shown in Figure 4.

1.3.3.1 Immediate Reporting

Prior to first and second phase readout of satellite and aircraft missions, there are a number of activities performed by the IIS as soon as data is available about mission coverage. The purpose of these mission readiness activities is to prepare for an efficient readout. Included are the preparation of:

a) Prediction of installations to be covered by a mission. These are sorted into two lists, one with targets in order of priority for first phase readout, and the other with targets in order by pass/frame for second phase readout.

b) Lists of targets to be assigned to PI readout teams based on team area specialties.

c) Readout workload estimates of each team to aid in team assignment of PI's.

d) Predicted film coordinates of targets to supplement data in the IDF for first phase readout, and to eliminate the plotting of coverage of each frame on $1:10^6$ scale maps.

e) Instructions for film distribution and control.

f) Presentation sheets for predicted targets.

Mission readiness activities differ from the current operations in that an automated target prediction scheme replaces the manual method used on satellite missions in support of second phase readout. The method is also used in support of first phase satellite readout, making the Center independent of prediction currently done elsewhere, and to support aircraft mission readout.

Readout for Immediate Reporting proceeds using the film, Presentation sheets, and supplementary support material maintained locally within the PI

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

TOP SECRET

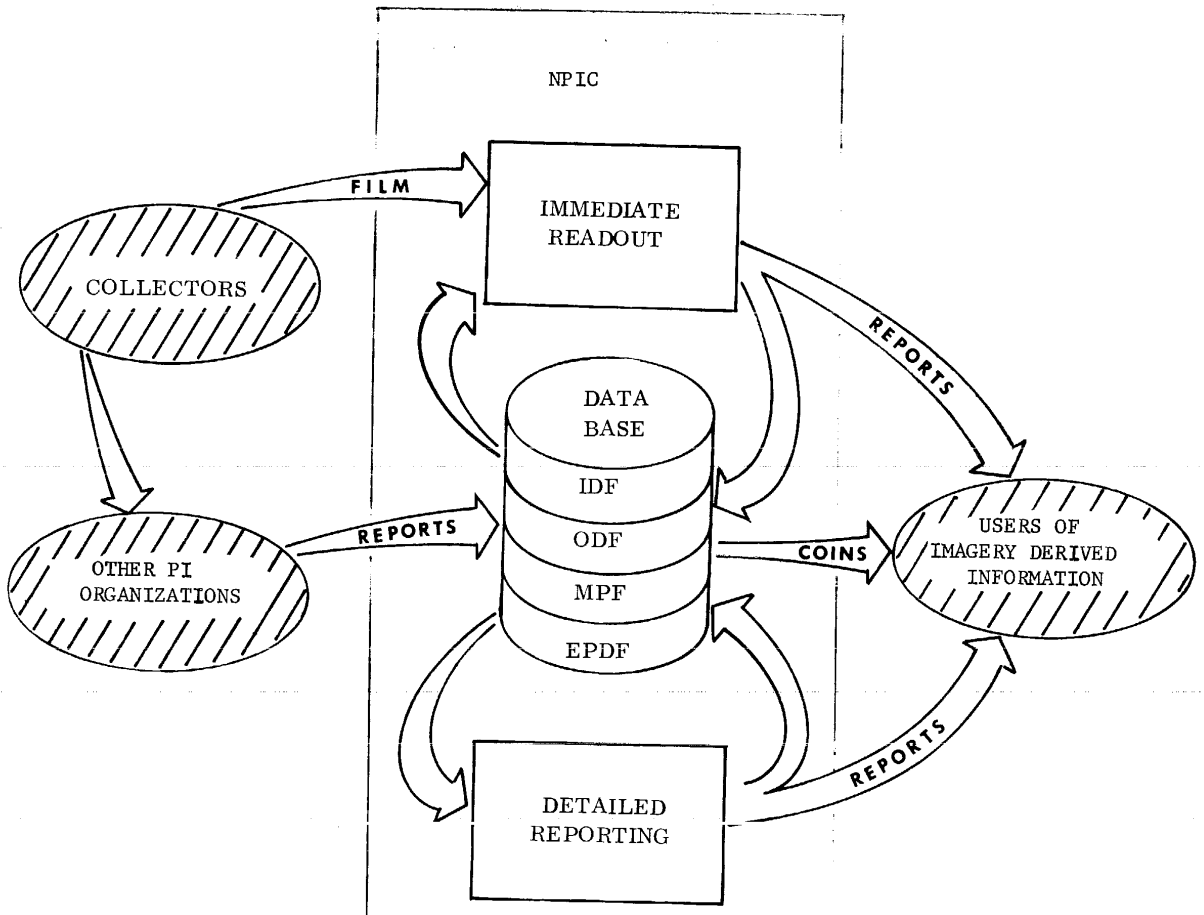


Figure 4 Role of NPIC In The Community

25X1

TOP SECRET

25X1

SCR 352

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

work area. Since a given target or geographical area is generally read out by the same PI team, the supplementary support material does not have to be moved about. The PI can also query the data base to request additional data during readout. On-line mensuration is available although, during these phases, the data in the Mensuration Parameters File for some camera systems will not have sufficient accuracy for all the types of measurements that might be desired. As readout is completed, the PI fills in a pre-printed Worksheet which organizes data in proper form for entry into the IIS. Provision is also made for the PI to recommend cable reports, briefing board candidates, and to identify new targets.

The results of readout are passed through an entry, verify, edit, and approve procedure to update the data base and to generate NPIC products. The objective has been to get data into machine readable form early in the cycle and to make readout data quickly available to the community. For these reasons, the individual readouts are handled separately. The text for cables is available to the Communications Section immediately. The readout data for reports (cable and hard copy) are batched on a daily basis, sorted into publication order (sterilized and unsterilized versions), and used to prepare mats for printing and the output to communication circuits. Provision is also made for providing selected portions of the readout data, such as OB data, available to other agencies in machine readable form for data base update purposes (via communication circuits or magnetic tape).

1.3.3.2 Detailed Reporting

Detailed reporting, or Third Phase readout, is a retrospective, in depth study of an installation, group of installations or other topic. It differs from Immediate Readout in the time devoted to the readout and the time scale of interest. Materials pertinent to Third Phase reporting, i.e. photography, reports, and other collateral data to be considered are identified by searching the data base, using the various indices available within the IIS. Actual readout of photography is again facilitated, as for Immediate Reporting, by a query capability to obtain supplementary information from the data base. Further aid to the PI is provided through an ability to gather, correlate, and summarize data on a multi-target basis from the data base. On-line data entry and editing are employed for updating Basic Intelligence Support Exploitation Reports and for extracting information from the data base for inclusion in reports.

1.3.3.3 National Data Base Activities

As previously noted, the files of the National Data Base of Imagery-Derived Information which NPIC is tasked with maintaining are the Installations Data File, Moving Target File, Objects Data File, Mensuration Parameters File, and Exploitation Products Data File. Inputs to these

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

files come both from within NPIC as a result of direct PI readout and from information generated by other agencies. Data base activities consist of file maintenance and the query capability.

In regard to file maintenance, all photo interpretive data resulting from NPIC Immediate and Detailed reporting is immediately entered, upon approval, into the Data Base and properly indexed by an automatic on-line updating procedure. All other material which will be used to update the data base is entered when acquired, and the automatic update sequence initiated.

The query capability is the means by which the user of the IIS can extract information from the data base. The users of the IIS include personnel both within, and external to, the Center. Users in the Center who frequently require information from a limited portion of the data base can achieve a limited proficiency in exploiting the data base with a minimum amount of training. A limited number of "Information Specialists" will be available who are proficient in exploiting all aspects of the data base. External users will communicate with the IIS through COINS, or will interface with the Information Specialists in person or by telephone. Center user interface with the system is through consoles. External users may generate the query language statements directly or have them generated by the special interrogation programs being developed as part of COINS I.

The query language, "PIRL" (PIC Information Retrieval Language), is designed to provide the following capabilities with a response time on the order of seconds:

a) The ability to retrieve sectors of a unit record by stating the file, unit record, and sector.

b) The ability to retrieve one or more fields of a sector by stating the file, unit record, sector and field(s).

c) The ability to retrieve fields of a designated file, unit record, and sector on the basis of two qualifiers within the fields (category type and Time or Mission Number).

d) The ability to establish the set of unit records containing designated terms in one or more classes such as Country, Military District, Portion of an alphanumeric classification list, Geographic location, and Object designators.

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

e) The ability to indicate to the originator of the query, the size of the set established by (d) before providing (only on request) a list of the set itself.

f) Provision of a dialog capability to further refine the set established by d) and e) by adding or deleting to expand or restrict the established set.

1.4 Report Organization

In the sections of this volume of the report which follow, the requirements on the IIS are established by describing the role of NPIC within the intelligence community (Section 2) and the workloads that the Center must handle (Section 3). Next the overall operation of the system is described in terms of the system configuration, the data base organization, and the activities of the Center (Section 4). (The details of the system design and operation are given in Volume 2 of this report.). Section 5 presents the major design studies conducted and the alternatives considered as part of the detailed design effort. (Additional studies are described in Tabs a through o, Part III of this report.) The final section provides a cost/effectiveness evaluation of the IIS in comparison with an extrapolation of the current Center operations to handle the anticipated workloads.

Details of the analytical methods used in the design studies reported in Section 5 are presented as Appendices.

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

2. ROLE OF THE CENTER WITHIN THE COMMUNITY

A major purpose of the IIS is to enhance the capability of NPIC to carry out its role within the intelligence community. It is therefore worthwhile to review this role as background for a discussion of the IIS itself. The responsibilities of NPIC which will be affected by the IIS can be considered in the following three areas:

a. Immediate readout of certain photographic reconnaissance missions.

b. Detailed reporting on certain categories of installations.

c. Maintenance of the four parts of the National Data Base of Imagery Derived Information (NDBIDI) concerned with installations, objects, exploitation products, and data required for mensuration.

These responsibilities are summarized below and their relationship to the rest of the intelligence community are shown on Figure 4.

2.1 Immediate Readout

Actual readout starts with the receipt of film from the collectors, although on satellite missions and some aircraft missions data describing the mission coverage is received prior to the film. This makes it possible to gather together, prior to the receipt of film, certain but far from all of the material used by the PI during readout. Currently the data describing mission coverage is received before the film is delivered to the processing site. This gives ample time to prepare for readout. However, improvements in the processing cycle or the advent of direct readout could reduce this preparation time significantly. Further, there are a substantial number of aircraft missions for which data about mission coverage is not available in advance. These factors make it desirable to reduce the amount of time required to prepare for a mission.

Immediate readout takes place in two modes referred to as "first phase" and "second phase". During the first phase a number of specific targets are read out in accordance with a set of readout requirements. During the second phase, the film is scanned to look for new activity and to index all known targets. In both cases, information from the data base is used to support the readout. The most significant information from these readouts is disseminated to the intelligence community through cables, briefing boards, and hard copy reports (e.g., OAK, IPIR, OAK Supplement). All of the derived data is used to update the data base.

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

While time is generally considered to be of utmost importance in the preparation of first phase reports, no specific requirement of general application can be established. Currently the time varies from one day to one week depending on the number of targets covered on the mission. This is considered compatible with the time required to collect and process the photography. Only a moderate increase in the amount of photographic coverage is expected in the future. However the amount of detail in the photography and the number of targets read out is expected to increase significantly.

2.2 Detailed Reporting

Detailed, or "third phase", reporting is an in-depth, retrospective investigation of a particular installation, group of installations, or other subject of intelligence interest, and may use imagery from several missions. The requirements for these investigations normally originate outside the Center and may be in response to the national needs or the needs of departments which do not have photo-interpretation capability. In carrying out these investigations, the data base is used to identify the available imagery and collateral data and to provide a summary of the information that has already been derived from imagery. Hard copy reports are issued and the newly derived information is used to update the data base.

Recent plans of the intelligence community, as described in the National Tasking Plan (NTP), call for a series of Basic Intelligence Support Exploitation Reports to be issued on a periodic basis. The Center's share of these reports represents a significant increase in work load over current detailed reporting activities.

2.3 National Data Base Responsibilities

It is in this area that the Center faces the most radical departure from current operations. Currently the Center maintains a data base primarily to support its own activities. The major component is the Target Brief File which is installations oriented and contains data from the readouts conducted by the Center. Copies of this file are distributed periodically to several members of the intelligence community. In addition, an All-Source Index to photo-interpretation reports prepared by all members of the intelligence community is compiled and distributed.

In the future the Center must maintain an Installations Data File (IDF) that fulfills national needs and contains readout data from all agencies in the intelligence community rather than NPIC readout data only. An Objects Data File (ODF) must also be defined and maintained. In addition the Center must maintain an Exploitation Products Data File (EPDF), similar to the All-Source Index, and a Mensuration Parameters File (MPF) which contains the mission and camera data necessary to derive measurements from the photography.

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

The intelligence community will gain access to this data base through the Community On-Line Intelligence System (COINS) or similar network established to handle the compartmented imagery derived information. While no requirement has been established on the maximum time to respond to a query, it follows that the value of this service increases as the response time decreases. In fact, as with most information retrieval systems, the value increases sharply as the response time is reduced to the point where the searcher can carry on a dialog with the data base and be informed as to the volume of data identified by the query before the information is delivered. This capability eliminates the need to have a highly refined query to start with; the query can be refined as a result of the responses received. In addition, and this is most important, the searcher is able to probe the data base while maintaining his train of thought and discover things that might not otherwise be found. Achievement of these benefits requires a response time on the order of five seconds.

2.4 Summary of Major Changes

While a large increase in the volume of reconnaissance imagery is not anticipated, it may well become necessary to reduce the time required for exploitation in order to keep abreast of needed improvements being sought in the collection and processing activities. Therefore, aids to reduce this time should be considered.

The advent of COINS marks the beginning of an important trend that will have great impact on the operation of the Center. It is our feeling that, as the intelligence community becomes used to COINS, there will be less and less dependence on hard copy reports for current information and more and more dependence on querying the data base. Therefore it becomes mandatory that the Center maintain its share of the national data base in a form that is convenient for external users. The implication is strong that the bulk of this data be held in random access storage in order to provide rapid response to queries. In addition, consideration should be given to making readout information available in machine readable form for use by other agencies that may have need to incorporate imagery derived information in their own data bases.

The most convenient form for detailed information will continue to be the hard copy report. However, the set of these reports covering the installations and subjects of interest to the community must be maintained on an up-to-date basis. Therefore provision must be made to issue or modify these reports on a periodic or as-required basis.

3. INPUT-OUTPUT LOADING FACTORS

3.1 Introduction

3.1.1 Purpose

25X1A

This section presents the primary design parameters which [redacted] has used: 25X1A

a. As the design point for the Integrated Information System (IIS)

b. As the set of conditions under which the IIS is compared with an extrapolation of the current procedures for purposes of evaluation.

It is important to note that these parameters are an estimate of the normal level at which the system will probably operate during the 1969-1974 time period; they do not represent a capacity limit. If actual loads exceed those defined by these parameters, the modular design of the system simplifies expansion. Conversely, if loads do not reach the expected levels, then procurement of some equipment can be deferred with corresponding reductions in cost.

3.1.2 Data Sources

Wherever possible, the quantitative values presented in this section have been drawn from authoritative sources, e.g., The National Tasking Plan and other annexes to the JIIRG Report. In other cases, there is no body of information upon which to draw and the quantitative values are necessarily engineering estimates. An example of this is the number of COINS queries per day which might be expected. A third class of values is exemplified by the number of entries in the Objects Data File where there is a wide divergence of opinion among members of the intelligence community. Since we have no rational basis on which to accept one opinion as more authoritative than another, we have used a compromise figure. In all cases, the values given herein have been submitted to and received the concurrence of the Center.

3.1.3 Organization

In the following sections we shall present the values of the primary design parameters which have been used in the system design. These parameters are divided into four classes related to the three principal missions of NPIC.

25X1 [redacted] 25X1
TOP SECRET
Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3 [redacted]

a. Immediate Reporting
b. Detailed Reporting
c. Data Base Size
d. Data Base Activity

For each parameter or group of parameters, we have discussed the purpose for which they were used and the data source or engineering rationale on which they are based.

3.2 Parameters Related to Immediate Reporting

Table II summarizes the mission related parameters which affect the workload of NPIC during the immediate readout cycle.

Four classes of imagery collection missions are considered:

a. Orbital Search missions exemplified by the KH-4 [redacted] 25X1A
25X1A [redacted]

b. Orbital Spotting missions exemplified by the [redacted] 25X1A
[redacted] 25X1A

c. Aircraft missions for which NPIC has responsibility for both specific priority target readout and mission indexing including the reporting of new targets and change and OB data on existing targets.

d. Aircraft missions on which the immediate readout of priority targets is done by other agencies but for which NPIC has responsibility for area scan and other Phase II activity.

For each mission type a normal and maximum value of each parameter is provided where applicable. These two numbers are self-explanatory except in the case of "Number of Missions per Year". In this case, the maximum figure is used only as a measure of the maximum arrival rate of new imagery or, inversely, as a measure of the minimum interval between successive missions of the same type. This topic is treated more fully in Section 3.2.1 below. 25X1A

[redacted]

25X1
Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3 [redacted]
TOP SECRET

[redacted]

3.2.1.1 Purpose

These data were used:

a. To determine the total number of immediate readout reports which will be generated annually by NPIC.

b. To establish the time span within which immediate readout of mission film must be completed to avoid system saturation.

c. To determine the volume of information in the Mensuration Parameters File.

3.2.1.2 Sources of Data

a. The data on orbital vehicles were extracted from the JIIRG report and adjusted in accordance with the equivalency factor discussed in Section 3.2.1.

b. The data on "Aircraft - Phase I and II" were extracted from the JIIRG report.

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

c. The data on "Aircraft - Phase II Only" are estimated on the basis that one-third to one-fourth of the aircraft missions estimated in the NPIC 5-year plan of 1966 will be scanned for Phase II reporting by NPIC.

3.2.2 Feet of Film/Mission

3.2.2.1 Purpose

This information was used to assist in determining the time required to complete Phase I and Phase II readout.

3.2.2.2 Data Sources

This information was supplied by TID.

3.2.3 Targets Covered/Mission

3.2.3.1 Purpose

This information was used to assist in determining the workload on the Center during the immediate reporting cycle in terms of:

a. The number of individual target reports which must be processed through data entry, editing, approval and update for each mission.

b. The number of presentation sheets for the PI which must be prepared for each mission.

c. The number of PI queries of the data base during immediate readout.

d. The number of pages in the First Phase Report and the First Phase Supplement.

3.2.3.2 Data Sources

The data for orbital missions were obtained by counting the actual number of targets indexed on typical KH-4 and KH-7 missions and increasing the numbers obtained to take into account the growth of the total target base and the probable fragmentation of targets required because of higher resolution photography. Although somewhat higher than JIIRG estimates, they are slightly smaller than estimates developed by OSD based on an average of several past missions extrapolated in accordance with the anticipated growth of the total target base.

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

The division of COMIREX targets into two priority groups (high priority=1A and 1B, low priority=1C) is a simple 1:3 ratio corresponding to the ratio for the total target base.

Data for aircraft missions is derived from the JIIRG report.

3.3 Parameters Related to Basic Intelligence Support Activities

Table III shows the expected normal and maximum values of several parameters affecting IIS workloads in support of detailed (Third Phase) reporting activities.

3.3.1 Purpose

These data were used to assist in determining:

a. The number of detailed reports, both in long and brief forms, that must be prepared by NPIC per year.

b. The number of different pages of material in support of basic intelligence activities which must be edited and published by NPIC per year.

c. The number of queries to be levied against the data base in support of basic intelligence activities.

d. The number of updates to the basic target description in the Installation Data File which will result from basic intelligence support activities.

3.3.2 Sources of Data

These data were extracted from the National Tasking Plan (rounded to hundreds). The normal is that provided in the tables for the year 1971. The maximum value is as given in the report.

3.4 Parameters Related to the Data Base Size

Table IV provides basic information on the numbers of unit records to be stored in the Installations Data File, the Objects Data File and the Exploitation Products Data File. The size of the fourth principal file for which NPIC has responsibility (the Mensuration Parameters File) is not shown on this table since the required information can be derived from the number of missions shown in Table II.

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

3.4.1 Purpose

This information was used to assist in establishing:

a. The quantity and types of mass storage media required to store the total data base.

b. The response time to queries against various files.

3.4.2 Sources of Data

a. Installations Data File

These values are derived from the JIIRG report. It should be noted that the NIETB total includes the active COMIREX targets and the NPIC total adds 6000 targets (which NPIC maintains for its own purposes) to the NIETB total.

b. Objects Data File

The total number of objects had been selected as 250 based on the number of objects described in the "DIA Fact Book of Communist World Forces" which covers military equipment down to the level of 60 mm mortars. We recognized that certain classes of industrial, transportation and agricultural equipment may also be of importance but have no information on which to estimate quantities. This number was increased to 1000 by the Center with the following qualification: "An object (as currently envisioned) is in effect an object class e.g. an object might be an 'aircraft'. The hierarchical structure within this object class would permit storage on specific aircraft types and the location of these types."

We have tacitly assumed that if an object is worth recording at all, one item of information required will be the locations at which it has been seen. The number of different locations will obviously vary widely for different classes of objects, however, we have estimated that 250 different locations represent a mean for a small number of object types, say 1000.

c. Exploitation Products Data File

This file will continuously increase in size. The data provided are based on a linear extension of current values.

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

3.5 Data Base Activity

Table V presents values related to the number of transactions against the data base on a daily cycle exclusive of the internal queries related to immediate and detailed reporting.

3.5.1 Purpose

The values presented in Table V were used to assist in establishing:

a. The response time to queries.

b. The currency of information contained in the data base.

3.5.2 Sources of Data

a. Queries

There is no body of experience on which to estimate how frequently outside users will query the NPIC data base. However, since experience with other systems (e.g. the one at NSA) has shown that once such systems are established, the frequency of use tends to grow rapidly, we have chosen to assume that there will be a total of 1000 users in the various members of COINS who will average one query per day against the NPIC data base. The maximum number of such queries in any one day is twice this number or 2000.

Queries will generally be answered by 2000 BCD characters of data. However, if a complete unit record from the Installations Data File were requested, up to 50,000 characters would be generated in reply.

b. Updates

The remaining values in Table V represent updating transactions against the files.

Since NPIC will maintain a National Data Base it will be necessary to add to the file PI reports prepared by external sources. If one-half of the reports currently received were added to the file, the average daily update rate would be 60 new reports.

Changes in requirements and significant new collateral items added to the file are estimated at one per active COMIREX target per year or approximately 25 per working day.

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

The present figure is approximately 100 per mission. Maximum values are estimated at twice the normal value.

If the target base is to grow from 21,000 to 36,000 over a five year period then the average growth rate is sixteen targets per day. However, since new targets can be expected to be added specifically as the result of missions, they will tend to be added in "batches" and the maximum rate is estimated at three times the normal.

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

4. OVERALL SYSTEM OPERATION

It is the purpose of this section to give an overview of the operation of the IIS. The rationale for choosing the operation described is given in Section 5 and in Tabs a through o in Part III of this report. The details of the design and operation are given in Part I, Volume 2 of this report.

The overall operation of the IIS is described in terms of the three activities of the Center that are affected, i.e., Immediate Readout, Detailed Reporting, and National Data Base Activities. A summary of the IIS configuration and the data base organization is given first to provide the background for the discussion of the activities.

4.1 System Configuration

The system configuration is described below in terms of the central computer complex, the input/output devices of the system, and the software provided to carry out the operations of the system.

4.1.1 Computer Complex

The computer hardware and related components which must be combined to support the IIS are shown in block diagram form in Figure 1. The equipment shown in this configuration has been influenced by several factors:

a. By agreement, Dual Univac 494's with their associated peripherals will be used as the central computer complex. (The cost of converting existing capabilities from the 494 system would outweigh possible economies in hardware over the life of the IIS. The dual 494 systems have sufficient capacity to handle all requirements of the IIS).

b. Existing remote equipment which is already owned by NPIC must be used to the maximum extent possible.

c. In the selection of new equipment, commonality to reduce costs and maintenance problems and the use of equipment with multiple capabilities to improve utilization has been a principal design consideration.

d. The physical separation of various NPIC divisions and branches utilizing comparable equipment has made duplication of some hardware necessary where actual utilization does not warrant it.

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3
25X1

With these factors in mind, the hardware configuration pictured in Figure 1 was selected to meet the operational capabilities of the Center in a time frame as of 1971. It should be noted here, that this hardware configuration will develop in an evolutionary sense from the present NPIC hardware configuration. A subset of the present NPIC hardware will be utilized for the initial IIS implementation (late 1968). This initial hardware configuration is shown in Figure 5. The evolution of hardware to the 1971 time frame, then becomes a simple matter of adding or deleting components from the initial hardware configuration. The details of the selection process that has led to the 1971 configuration are presented in Volume 2.

4.1.2 Consoles

The remote man/machine interface units and their distribution throughout NPIC are shown in Table VI. These devices represent the primary method by which Center personnel communicate with the IIS. By the use of these devices, data may be entered into the system, manipulated within the system and extracted from it. Since many of the devices serve multiple purposes, it is most convenient to describe the functions which must be served and then indicate the particular hardware units which are applicable.

4.1.2.1 Data Entry

There are a number of areas where new data must be entered into the system to update the files:

- On the completion of an individual target readout during Phase I and Phase II Readout
- On the confirmation of a new target
- When significant collateral information is available to update the IDF
- When readout requirements specified by external sources are modified
- When any of the internal files of the system (e.g. Film Location File) are to be updated
- When field reports from exploitation units outside of NPIC are received
- When indexing data concerning reports generated by exploitation units outside NPIC is available to update the EPDF.

25X1
Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

TOP SECRET

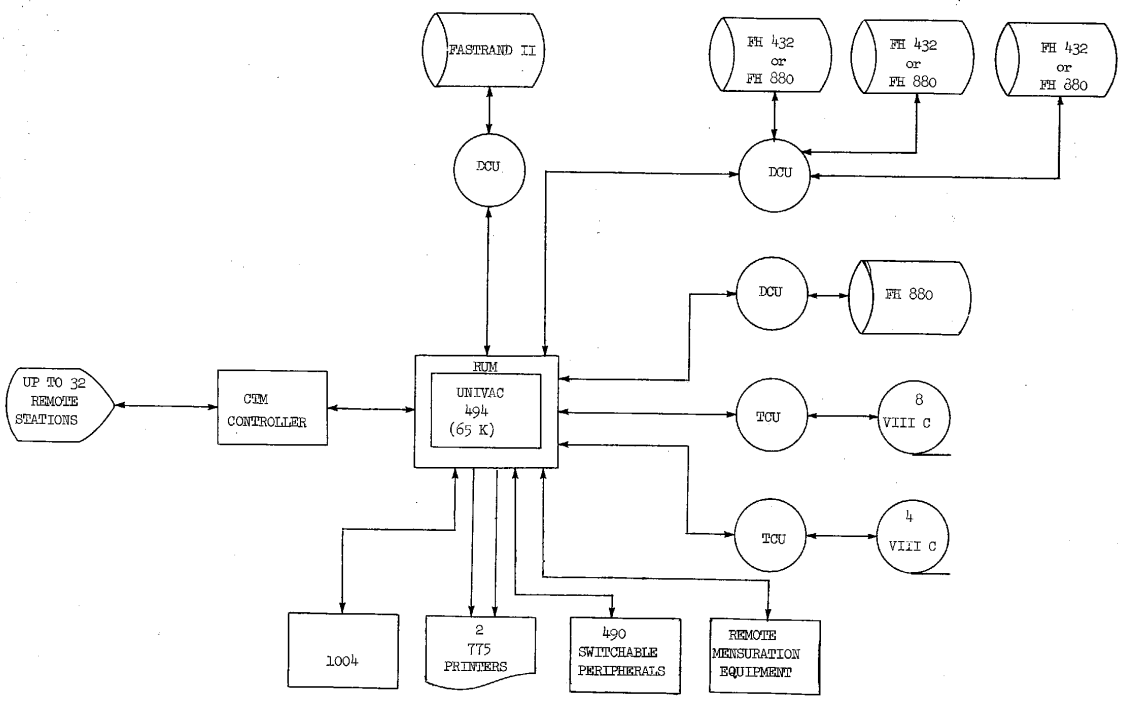


Figure 5 Initial IIS Hardware Configuration (1968)

25X1

TOP SECRET

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

- When new or revised Mission Correlation Data is received to update the MPF (or Position-Attitude data for KH-4 missions)

All of these operations except the last require that manuscript information be converted to machine-readable form. (Mission data is received in machine-readable form, either as magnetic tapes or punched cards from the 1004).

Wherever data must be manually entered, two factors must be considered:

a. A keyboard must be provided to enter the data initially.

b. A "read-back" display must be available so that an operator may visually verify that the data was entered correctly.

These capabilities will be provided in one of two ways:

a. NPIC component responsible for the entry of large volumes of data (e.g. CSD, PD) will be provided with a keyboard/CRT display unit with a self-contained buffer. The data is keyed in and immediately appears on the CRT. After visual verification (or correction if required) the contents of the buffer are transferred to the computer memory. One major data entry activity, individual target readouts resulting from First and Second Phase readout, is immediately followed by a separate Verification step. Hence these entry consoles do not require the display unit. It should be noted that if the data is to be used immediately to update a National Data Base File (i.e. it has been approved) the entry may be immediately followed by a query to make sure that no error was made in the transfer process.

b. Components which make infrequent or very brief entries (e.g. Charge Out Records in PSD) into the system will be provided with keyboard/hard copy displays of the Teletype variety. Since the entries made by these components modify only files which they maintain for their own purposes, the hard copy which is created during the keying process is adequate for verification. Corrections can be made immediately by re-punching the data. As in the first case, a query can be used to check the content of the file.

Within the IIS, queries may be handled in either of two ways:

a. The individual formulating the initial question may convert it to a "Query Statement" in the language called for by the system and key it into the system himself.

b. He may utilize the services of an Information Specialist who performs the conversion and key-in operations and returns the reply to the query originator, either as a display or hard copy.

It is anticipated that in the IIS, the first method will be used only when the data required can be retrieved by entry of a reference number or at most a simple and explicit query. The second method will be used for complex iterative queries.

The function may be served by either of two types of equipment:

a. A keyboard/CRT of the same type used for editing

b. A Teletype keyboard and medium speed printer

The former type of equipment will more generally be used, particularly by the Information Specialists and particularly when the query is such that a dialog with the data base must be carried on. It should be noted however that even though the actual query operation is carried out at one location, the reply may be directed toward any display by the operator. Thus a PI who has a complex question may telephone the Information Specialists, who after completion of the query cycle, directs the reply to a medium speed printer in the immediate vicinity of the telephone.

The Teletype keyboard and printer are used for queries which can be formulated as simple reference numbers or explicit statements and when a hard copy of the ultimate reply is required. More complex iterative queries could also be processed with these devices but the relatively low speed of the output printer (40 characters/second) is not particularly suitable for "dialog" operations.

4.1.2.3 Verification, Editing and Approval

When information which is derived from imagery within the Center is entered into the IIS, it must be verified, edited and approved before it is published or becomes a part of the National Data Base. This is accomplished by the operator calling up the text for correction from the computer memory. Changes to the display are made by pointing to the location of the change with the light pen or cursor, making the change on the keyboard and depressing the appropriate function key. The corrected text is then

25X1 [redacted] TOP SECRET Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3 [redacted] 25X1

transmitted back to the computer memory. These units must have upper/lower case capability. It should be noted that formatting can be accomplished by the insertion or deletion of blank characters.

4.1.2.4 Plotters

A variety of computer driven or magnetic tape driven plotters will be used in the system to provide graphic outputs. Data to be plotted may be the result of mensuration in the form of site plots or engineering drawings, the result of analytical computations, or plots of mission coverage data. These plotters are already owned by NPIC and no new acquisitions are planned.

4.1.2.5 Miscellaneous Man/Machine Interface Devices

In addition to the specialized equipment described in the previous sections, the IIS will use several standard computer input/output units to serve as man/machine interfaces. Included in this set are:

a. A card reader for input of KH-4 mission data, new programs and data for "batch" computations of all types.

b. A paper tape punch to generate the necessary input tape for automatic photo-typesetting.

c. A high-speed (600-1000 lines/minute, upper/lower case), printer to generate large volume output from the data base, in particular the PI Presentation for Phase I and Phase II Readouts and the mats for the First Phase Report and Supplement.

d. A Teletype automatic send/receive unit (ASR) for the preparation of the paper tape and verification copy of reports to be distributed as cables.

4.1.3 Software System

The success of the IIS as an improved information system is dependent in large measure on the computer programs which support and are, in fact, an integral part of the system. These computer programs are the vehicle which drives the IIS hardware and, in turn, provides the capabilities of the IIS. These computer programs are best described in terms of the system functions requiring computer support, and how the users of the system interface with the information (data base) content of the IIS by utilizing these computer programs.

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3 [redacted]
TOP SECRET

On a functional level the IIS software can be defined in terms of the support it provides to sixteen (16) functional areas. These areas are defined as follows:

a. File Update and Maintenance - Creation of sections of, addition of data to, and deletion of data from the operational data base (the ODF, IDF, etc.).

b. Data Entry - Acceptance of input data from various equipment, (keyboard, card reader, etc.), error checking, conversion and formatting of the data for further processing.

c. Verification, Editing and Approval - Support the VEA (Verify, Edit, and Approval) function, including the generation, modification and regeneration of displays to consoles manned by verifiers, editors, or the approving authority.

d. On-Line Mensuration - Support to real-time mensuration for PAG, TID, or resident detachments.

e. Query Handling - Decoding of a user-oriented Query Language and retrieval of desired sets of information from the operational data base (ODF, IDF, MPF, EPDF and other files).

f. Cables and Real-Time Reports - Formatting and preparation for transmission of either computer-to-computer information or outgoing cables.

g. Report Generators - Generation of hard copy reports automatically from information stored in the data base.

h. Executive - Control of the entire IIS software system, including all I/O handling, program scheduling, security, maintenance, etc.

i. Update Mensuration Parameter File - Generation of the necessary parameters, on a mission basis, to support both on- and off-line mensuration.

j. Batch Mensuration Support - Similar to on-line mensuration (d) except that support is provided by the IIS in an off-line sense generally using punched cards as input.

k. Target Prediction - Prediction of targets covered by a given mission prior to film arrival.

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

1. Assignment - Assignment of film and necessary support materials to various PI teams for Phase I and Phase II readout.

m. File Maintenance Language - Processing of an on-line input language similar to the Query Language (e). The language provides the user interface with the maintenance procedures themselves (a) e.g. CSD corrections to Header Information or the addition of collateral material to the IDF.

n. Presentation Language - Processing of an on-line language which permits the user to change or create report formats.

o. Miscellaneous Software - Provision of a collection of programs to support the photo reference file, maintenance and system operation.

Programs to support these sixteen functional areas may be logically grouped into eight (8) program packages (groups of functionally related programs). The eight program packages and the functional areas they support are:

1. Update Package

   a. Update and Maintenance

2. Data Entry Package

   b. Data Entry

3. VEA Package

   c. Verify, Edit, and Approval

4. Mensuration Package

   d. On-Line Mensuration
   j. Batch Mensuration Support

5. User Language Package

   e. Query Handling
   m. File Maintenance Language
   n. Presentation Language

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

6. Report Generation Package

g. Report Generators

7. Executive Package

h. Executive

8. Computation and Support Package

f. Cables and Real-Time Reports
i. Update Mensuration Parameter File
k. Prediction
l. Assignment
o. Miscellaneous Software

The relationship among these eight program packages can best be illustrated in terms of how they are employed when a user interfaces with the IIS data base. Figure 2 relates this user interface to the data base through the IIS software. The IIS data base is represented by the inner circle in Figure 2. Surrounding the data base circle are the eight program packages. The outermost segment of the figure represents the users. The arrows going to and from the users through the software to the data base picture the user interface. The counter clockwise arrow, originating in the program language and terminating in various other packages, represents the compiler function of generating the software in six of the eight program packages.

4.2 Data Base Organization

The heart of the IIS is the data base which supports the operations within the Center and provides information to other members of the intelligence community as well. The Center's portion of the National Data Base of Imagery Derived Information has been defined by the National Tasking Plan and consists of data on installations and objects, information about exploitation products (PI reports), and the information about the imagery that supports mensuration. The data maintained by the Center must include data developed by other agencies as well as data developed by the Center itself.

For ease of handling, the total data base has been divided in two separate ways. The first is the physical division which is concerned with allocating the various parts of the data base among the various storage media. This puts the most used data in the most accessible storage media and less used data in progressively less accessible storage. While certain assignments are made in this report, they should be considered tentative since the assignments can be changed in accordance with the experience gained during operation.

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

The second division is the logical division of the total data base into files, the files into unit records, and the unit records into sectors and further subdivisions. The major files consist of the Installations Data File (IDF), the Objects Data File (ODF), the Exploitation Products Data File (EPDF), the Mensuration Parameter File (MPF), and a fifth, which may be considered a subset of the IDF, the Moving Target File (MTF). Each file is divided into unit records which contain all the information about a given element, e.g. a unit record in the IDF contains all the information about a given installation. Each unit record is divided into sectors, one for each type of information about the element, e.g. one sector in each unit record in the IDF contains the header data, another the installation description, etc.

Access to these sectors is provided by an index record, one for each unit record. This index record gives the location of each sector by first giving the storage media, i.e., drum, tape, hard copy, and then the address within the media. These index records are kept on the FH1782 drum for easy access. Thus it can be seen that the data itself can be moved among the various storage media in accordance with usage experience provided that the necessary modifications are made to the index records.

Each of the five major files will now be briefly described in turn, followed by a description of the file indices that provide access to the unit records through the index records.

4.2.1 Installations Data File (IDF)

This file is analagous to the current Target Brief File. It will eventually contain one unit record for each of the 42,000 installations. Each of the unit records are divided into the following thirteen sectors:

Header
Brief Fixed Facility Description
Changes to Fixed Facility Description
Detailed Fixed Facility Description
Readout Requirements
Collateral Data
PI Notes
Photo References
Source Data References
Concise Installation History
Complete Installation History
Formatted Fixed Facilities
Formatted Mobile Equipment

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

Most of these sector names are self-explanatory, however, in certain cases, a few words of explanation are appropriate.

a. Descriptions for those installations covered by Basic Intelligence Support Exploitation Reports are taken from the complete description presented in the report and the abstract (when provided). These become the detailed and brief description respectively. The simple installations will have only a brief description. In either case the descriptions are the responsibility of the agency assigned third phase readout for the installation and the description in the IDF is only changed when a third phase readout has been accomplished by the responsible agency. If changes are noted during first or second phase readout, these changes are appended to the end of the basic description in the Brief Fixed Facility Description sector. Descriptions for those installations not covered by third phase reports will be derived initially from the current Target Brief File and validated or updated as required during each subsequent first or second phase readout.

b. Readout requirements currently exist in free text and can be used directly. However, if standard sets of readout requirements (EEI's) are established for various classes of installations, they can also be incorporated either directly or by reference in the index record. In addition, if such standards are established, the installation description could be formatted in accordance with them.

c. The Source Data References sector includes references to the most significant reports about the installations, other sources of collateral data, map references, and references to briefing boards about the installation.

d. The Concise Installation History starts with the negation date, first seen date, and the first description. This is followed by the chronological sequence of change statements that have been made about the installation. The purpose of this sector is to provide a concise description of how the installation has developed over the years.

e. The Complete Installation History, on the other hand, is a complete transcript of all that has been said about the installation and will be much like the current Target Brief File. Since it is retained principally for record purposes, it can be stored in hard copy and magnetic tape form rather than in an on-line storage media.

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3
25X1

The last two sectors contain numerical data formatted so that it may be summarized or otherwise operated on by computer methods. One is a tabulation of the fixed facilities at the installation. The other is a chronological listing of the mobile equipment that has been seen at the installation including the data and number of each type of equipment.

4.2.2 Moving Target File (MTF)

This file is designed to give the characteristics, and an historic record of the movement, of those unique mobile objects which are of primary interest to the intelligence community at any given time. In this context a unique object is one that a PI can distinguish as being one-of-a-kind, e.g., a specific atomic submarine identified by its registry number (rather than as a member of a class or type). These targets could be considered as installations and the data about them included in the IDF. However, there are enough differences in the types of data to be maintained and the uses to which it will be put to make a separate file worthwhile.

There is a unit record for each unique object such as an experimental ship or VIP aircraft and each unit record is divided into the following sectors:

Header
Dimensional Data
Description
Readout Requirements
Collateral Data
Photo References
Sighting Information

4.2.3 Objects Data File (ODF)

This file has been called for by the National Tasking Plan but has not yet been completely defined in terms of the requirements of the community. There is currently no body of digital data for incorporation in this file although the Center has been collecting descriptive material on a number of objects. It is planned to create a unit record for each unique type of object without being concerned with the question of "what is an object" nor the relationships among various types of objects. Each unit record contains the following sectors which represent the types of information that appear pertinent:

Header
Dimensional Data
Description
Photo References
Graphic References

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

Readout Requirements
Source Data References
Location Information

The header contains the information identifying the object, including such things as name, alphanumeric designator and manufacturer. In addition, the place of the object in an object classification scheme might be indicated as well as its relation to other objects, e.g. "next higher assembly". Dimensional data consists of standard parameters such as height, width, length, volume, and shape as well as provision for special parameters. If a set of standardized parameters that are sufficiently descriptive can be developed, the potential exists for searching this data as an aid in identifying unknown objects. Photo, Graphic, and Source Data References indicate the existence of aerial and ground photography, drawings, and other material descriptive of the object. As with installations data, standardized readout requirements can be accommodated and might also provide a format for the object description.

The location information sector contains a list of all the installations where the object has been seen. If detailed information, such as date and quantity is required, the appropriate unit records in the IDF can be consulted. This sector will provide the answer to questions such as, "Where have Widgits been seen?", without a serial search of the IDF. Since this data would not be useful for all objects, e.g. common trucks, the readout requirements will indicate whether or not it will be kept.

4.2.4 Exploitation Products Data File (EPDF)

This file is analogous to the All-Source Index currently maintained by the Center. A unit record is created for each report to be covered by the file. This unit record comprises the bibliographic record or "catalog card" for the report. It consists of the following sectors:

Header
Location Data
Alternate Titles

Again, the header contains the identifying information, i.e., issuing agency, agency report number, Center accession number, type of report, etc. The location data sector gives the geographical locations or place names described in the report. The alternate titles are generated to indicate the subject matter, content, and scope of the report.

This information is currently published periodically in hard copy form. Presumably the need for such distribution will decrease with the advent of an effective means of querying this file.

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

4.2.5 Mensuration Parameters File (MPF)

The Center has also been charged by the National Tasking Plan with the maintenance of a file of the parameters necessary to derive measurements from imagery. These parameters are of two types. The first describes the camera system itself and the second describes the position, attitude, and velocity of the camera and other derived data for each frame of imagery. Although the format of the data differs for each type of camera system the unit record concept is still used. Each unit record is a complete mission with the "header" sector giving the identifying data and the camera system constants and each additional sector providing the data for one pass. The data for each frame is a separately retrievable field within the pass sector. Provision is made for recording variances for all parameters.

4.2.6 File Indices

Rapid and efficient access to the files requires careful consideration of the method of gaining access to the index record for each unit record. If the elements in each file were always referred to by the same kind of term, e.g. WAC/PIC Number for the IDF, there would be no problem. The index records could be ordered in accordance with that term and access could be achieved without an extensive search. However, if access were requested on the basis of another term such as installation name, a serial search of all header sectors would be required to find the appropriate record. When the file is stored on a serially accessible medium such as tape this makes no difference, since the same amount of time is required to reach the appropriate record whether or not it must be searched for. However, serial search represents a great inefficiency with a randomly accessible storage medium such as a drum, and in fact, defeats the whole purpose of the random access memory. For this reason, the IIS has been provided with indices to relate all the commonly used access terms to the location of the index records of the corresponding unit records.

Two types of indices are provided, unique term and broad term. The unique term indices provide a one-to-one correspondence from the entry term to the unit record. This does not, however, rule out the possibility of having two terms (e.g. two installation names) that refer to the same unit record, nor does it require that there be a term in each index for every unit record. For example, there may be one index that relates terms in an object classification scheme to unit records in the ODF, but there may be objects in the ODF that do not fit into the classification scheme (though they may be indexed by name or manufacturer).

Broad term indices provide a one-to-many relation between terms such as a country name and all of the installations within that country. These are, in a sense, inverted files that eliminate the need to search a whole file to find the unit records which meet a certain criterion. No attempt

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

has been made to create indices for all possible search criteria, however, the ones that are provided can generally be used to limit the number of unit records that must be searched to find those meeting any given set of criteria.

Indices can be created on the basis of any of the terms that appear in the header sector or other parts of the unit records. However, the indices planned for initial implementation are shown on Figure 3 along with the files to which they provide access. The PAG and CSD Installations Files indicated on Figure 3 are hard copy and film chip files maintained by those two groups. The PAG files contain primarily mission photography maintained by the various readout teams. The CSD files include ground photography from various sources and cover many non-targeted installations.

4.3 Immediate Readout Activities

The Center performs immediate readout of satellite and certain aircraft missions in two phases, the first of which is target oriented, and the second of which is a systematic scan of the film. The IIS support to these activities is described below in terms of what is done to support preparation for readout, readout itself, and the activity required to make the readout data available to the community. First and second phase readout are described together since their similarities are greater than their differences.

4.3.1 Mission Readiness

There are a number of functions performed by the IIS as soon as data is available about the coverage of a mission, be it satellite or aircraft. The purpose of these functions is to provide Center personnel with the information necessary to plan for an efficient readout during both first and second phase. The results of these functions are shown on Figure 6.

The first step is to compare the predicted mission coverage with the geographic locations of the installations listed in IDF to determine those installations that will be covered by the mission. Although desirable, highly accurate mission data is not necessary for this function because the predicted camera coverage area can be modified to allow for inaccuracies in the position and attitude data. Next the list of predicted targets is sorted into two parts, one with the targets for first phase readout in priority order, and the other the targets for second phase readout sorted in pass/frame order. These lists are then matched against the target category and area specialties of the first and second phase readout teams respectively to produce a list of the targets assigned to each team. In addition to the list, the IIS estimates the workload of each team on the basis of the number of targets and the average readout time for targets of the type read out by the team. This information is used by Center personnel to plan the work assignments during mission readout. The list of predicted targets by teams is also used by CSD to prepare any support material that is

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

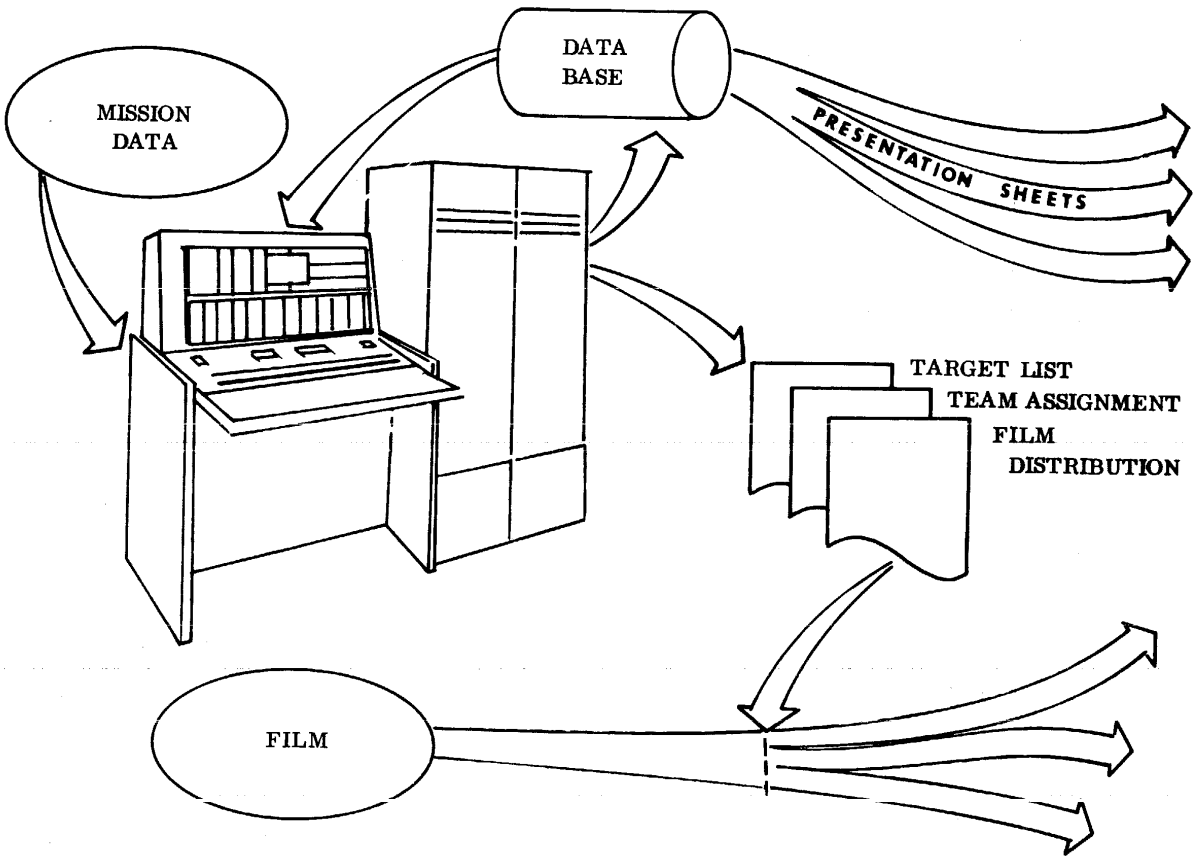


Figure 6 Mission Readiness Activities

TOP SECRET

25X1

TOP SECRET

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

not part of the IDF. If reliable cloud cover information becomes available at some time in the future, it can be used to eliminate cloud covered targets from the predicted target list and make the corresponding corrections to the team workloads.

The next step is to generate instructions for distributing the film among the PI teams. To obtain the most efficient use of the limited number of copies that are available, knowledge of the targets on each frame and the frames in each can is required. The former can be predicted with sufficient accuracy on many camera systems on the basis of the data on mission coverage (telemetry or flight data). However, on the KH-4 and some aircraft systems there is sufficient variation in the camera frame rate to introduce significant errors. Therefore provision is made to incorporate information from the processing site on the number of frames per camera operation to compensate for variations in camera frame rate. The film assignment function could treat each pass, or camera operation as one continuous piece of film. However if a pass is to be cut and placed in two or more cans at the processing site, this information could, in some cases, be used to develop a more efficient film assignment schedule. Therefore provision is also made to incorporate this information when it is available.

The result of the film assignment function is two sets of lists. One is organized for the personnel responsible for film control and provides the routing for each film can. The other is organized for the team captain and indicates the source and disposition of each film can his team will require. These two sets of lists should contribute significantly to the control of film during the immediate readout process.

An extension of the prediction of the targets that will be covered is the prediction of the film coordinates of the target. For first phase readout, this is based on the camera characteristics, the data describing mission coverage, and the information to compensate for variations in camera frame rate if available. After the film coordinates have been measured on some targets during first phase readout, the differences between predicted and actual coordinates can be used to refine the predictions on the remainder of the targets. This, in effect, compensates for errors in camera position and attitude data. For first phase readout, which is target oriented, the predicted film coordinates are added to the data that is withdrawn from the IDF for each target to be read out. For second phase readout, which is a systematic scan of the film, a separate list is generated which identifies each target on each frame (both active and inactive targets) and the corresponding film coordinates, and gives the geographic coordinates of the

25X1

TOP SECRET
Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

The final step is to extract the necessary data on each predicted target from the IDF and prepare the Presentation Sheets (described in Section 4.3.2) to support the PI during readout. These Presentation Sheets are printed out in accordance with the assignment of targets to PI teams, thus eliminating the hand sorting operation currently required.

The mission readiness activities outlined for the IIS differ from current operations in two major aspects. First, a computerized target prediction scheme is used to replace the manual method currently used on satellite missions to support second phase readout. The method is also used to support first phase satellite readout, making the Center independent of the prediction done elsewhere, and to support aircraft mission readout. The second, which is dependent on having the predicted targets and team assignments in machine-readable form, is the delivery of film and support materials sorted in appropriate order, directly to the appropriate PI teams. In both cases it is felt that there will be significant savings in manpower, and that the elapsed time required to prepare for the readout activities will be reduced. While this time saving is not important for satellite missions today, it is important for some aircraft missions, and may become important on satellite missions in the future. In addition, the automated procedures provide information which allows better control of activities during the readout process itself and identifies more completely the support materials required.

4.3.2 Readout Activity

The readout activity for both first and second phase is shown functionally on Figure 7. The material available to the PI includes the film to be read out in accordance with the distribution schedule developed as part of the mission readiness activities, Presentation Sheets containing information from the IDF on the targets to be read out, and supplementary support material that is maintained locally within the PI work area. Since a given target or geographical area is generally read out by the same PI team, this supplementary support material does not have to be moved about. The PI also has the ability to request additional data from the data base during read out. Although not shown in Figure 7, the IIS also provides for on-line mensuration during first and second phase readout. (However, during these phases, the data in the Mensuration Parameters File for some camera systems will not have sufficient accuracy for all the types of measurements that might be desired). As readout is completed, the PI fills in a Worksheet which organizes data in the proper form for entry into the IIS. Provision is also made for the PI to recommend cable reports [illegible]

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

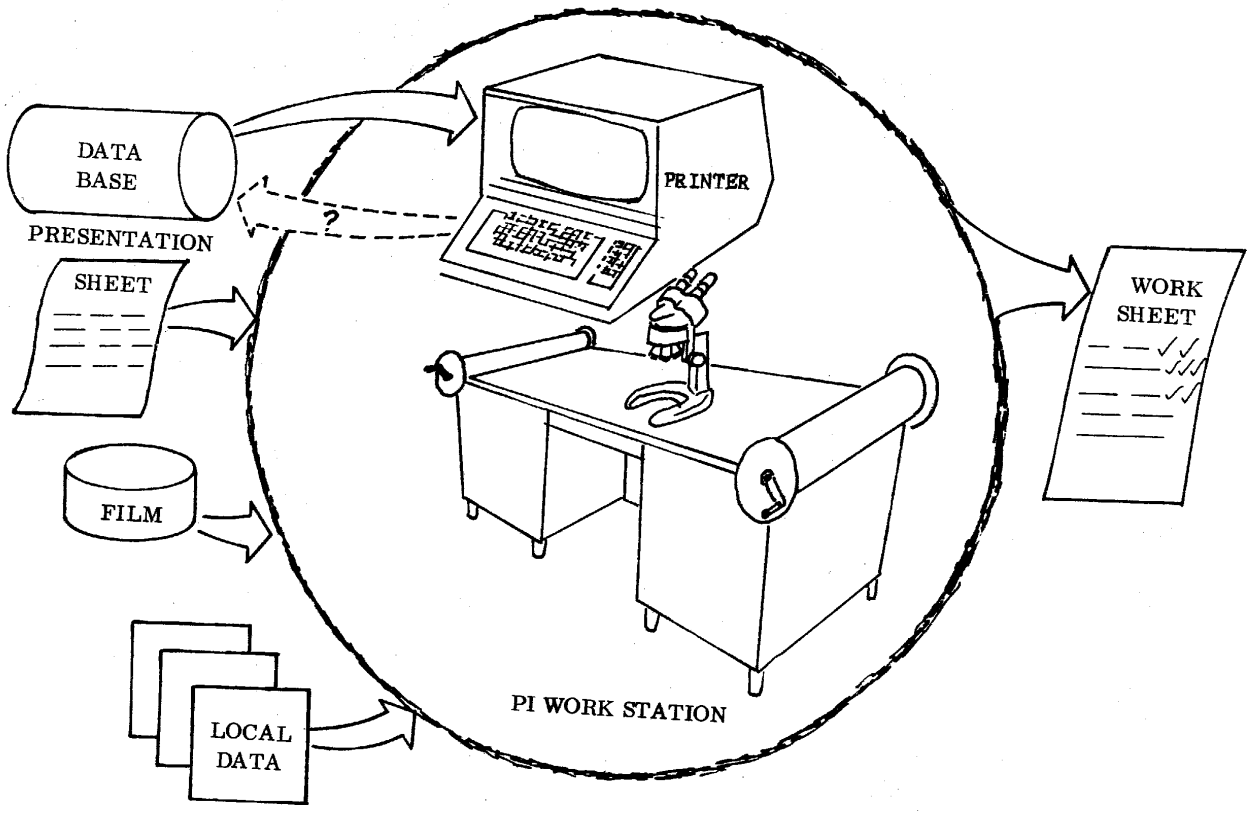


Figure 7 Immediate Readout

TOP SECRET

TOP SECRET

25X1

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

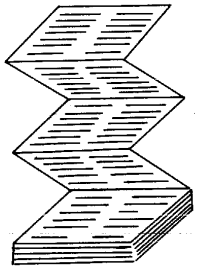
TOP SECRET

CURRENT SUPPORT

IIS OBJECTIVES

PROPOSED SUPPORT

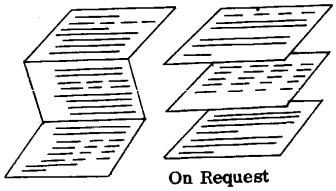
PRESENTATION - 1st AND 2nd PHASE READOUT



Standard Display

(Standard Display)

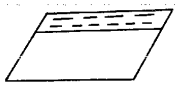
Elimination of extraneous data
Reduction in redundancy
Tailoring of readout support



Standard Display

On Request Displays

WORKSHEET FOR DATA RECORDING



Input data formatting for:
- Automatic product generation
- Automatic query response
- Automatic data presentation

Provision for PI data retrieval
PI control of presentations

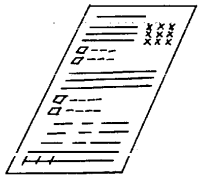


Figure 8 Development Objectives - Presentation and Worksheet

25X1

TOP SECRET

SCR 352

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

first by reorganizing the data currently contained in the Target Brief File into blocks in accordance with anticipated use. This reorganization (see the discussion of the IDF, Section 4.2.1) also reduces the volume of data by eliminating much of the redundancy that exists in the current Target Brief File. Second, the volume of data routinely given to the PI is further reduced by not including on the Presentation Sheet data that is seldom used on a readout. All data is, however, available on request from the data base. The data that is provided on the Presentation Sheet consist of:

Installation header
New or significant collateral
Readout requirements
Brief installation description
PI notes from previous readouts
Latest briefing board
Recent mobile equipment readouts
Recent photo coverage
Map references

An analysis (Section 5.2) indicates that this data represents about 13% of the data in the IDF for the average installation, but is sufficient to accomplish readout 80% of the time.

On first phase readout, the Presentation Sheet also contains the pass, frame, and predicted coordinates for all coverage of the installation. On second phase readout, which is a systematic scan of the film, a separate list is provided which identifies all installations listed in the IDF in readout order and gives their predicted film coordinates. Installations that have been read out in first phase and inactive installations (no readout required) are so indicated. The list also contains the geographic coordinates of the four corners of each frame.

The IIS Worksheet is highly formatted, providing spaces for the PI to fill in the appropriate information. For some types of information, multiple choice statements are provided. This achieves two objectives. First, it reduces the amount of text that a PI must compose and second, it forces the readout data into a standard format so that the various information elements can be utilized by the IIS without further interpretation.

Within the scope of the IIS, the major aid to the intellectual process of the PI during readout is better organization of the support materials. This has been done through organization of the Presentation Sheet and making the whole IIS data base available to the PI during readout. Although the exact nature of future PI aids such as change detection and target recognition equipment is not known, it is likely that they will require quantitative measures of target characteristics (e.g. dimensions, counts, angles, etc.). This information can be stored in the IIS data base and made available on

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

request, much as information is made available to the PI. Another potential benefit is that the film coordinate prediction procedure, as refined by film coordinate measurements during first phase readout, may provide sufficient accuracy so that film coordinates do not have to be read during second phase readout.

Even without these future possibilities, the IIS provides certain tangible benefits for the PI portion of immediate readout. These include better organization of support materials, accommodation of on-line mensuration, better control of film, and reduced report composition by PI's. In addition, the estimates of manhour requirements by PI team made as part of the mission readiness activities should improve the utilization of PI's during immediate readout.

4.3.3 Output Cycle

The output cycle is concerned with taking the results of immediate readout, passing them through a data entry, verification, editing, and approval procedure and utilizing the approved information to update the data base and generate NPIC products. The objective has been to get the data in machine readable form as early in the cycle as possible and to make the readout data available to the community as rapidly as possible. For these reasons the individual readouts are handled separately. The same cycle is also used for the text of cables initiated as a result of unusual activity discovered during readout.

When the Worksheet has been completed by the PI and approved by his team captain it enters the output cycle shown on Figure 9. In the first step the data from the Worksheet is keyed into an entry console, and enters a Working File which is in the random-access storage. This step also initiates a transfer of data that might be required in the output products from the data base to the Working File. This includes, as a minimum, the header information and might also include a negation of the readout requirements (e.g. "No new construction activity was seen.") or an installation description if called for by the Worksheet. The Worksheet is then passed to the verifier as an indication that the readout is available for review and as a reference for determining errors of omission during the entry process. He calls for it to be displayed on his console and makes whatever corrections are necessary using the data base as a reference as required. It should be noted that much of the verification currently required will not be needed with the IIS because the information is taken directly from the data base (e.g. place names, coordinates, etc.). When he is through he enters a completion statement to indicate that the readout is ready for editing. The editor retrieves the readout by calling for the "next readout", makes the necessary corrections, and enters a completion statement to indicate that the readout is ready for approval. At the completion of each step, the Queue List (index to the Working File) is updated to show that a

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

TOP SECRET

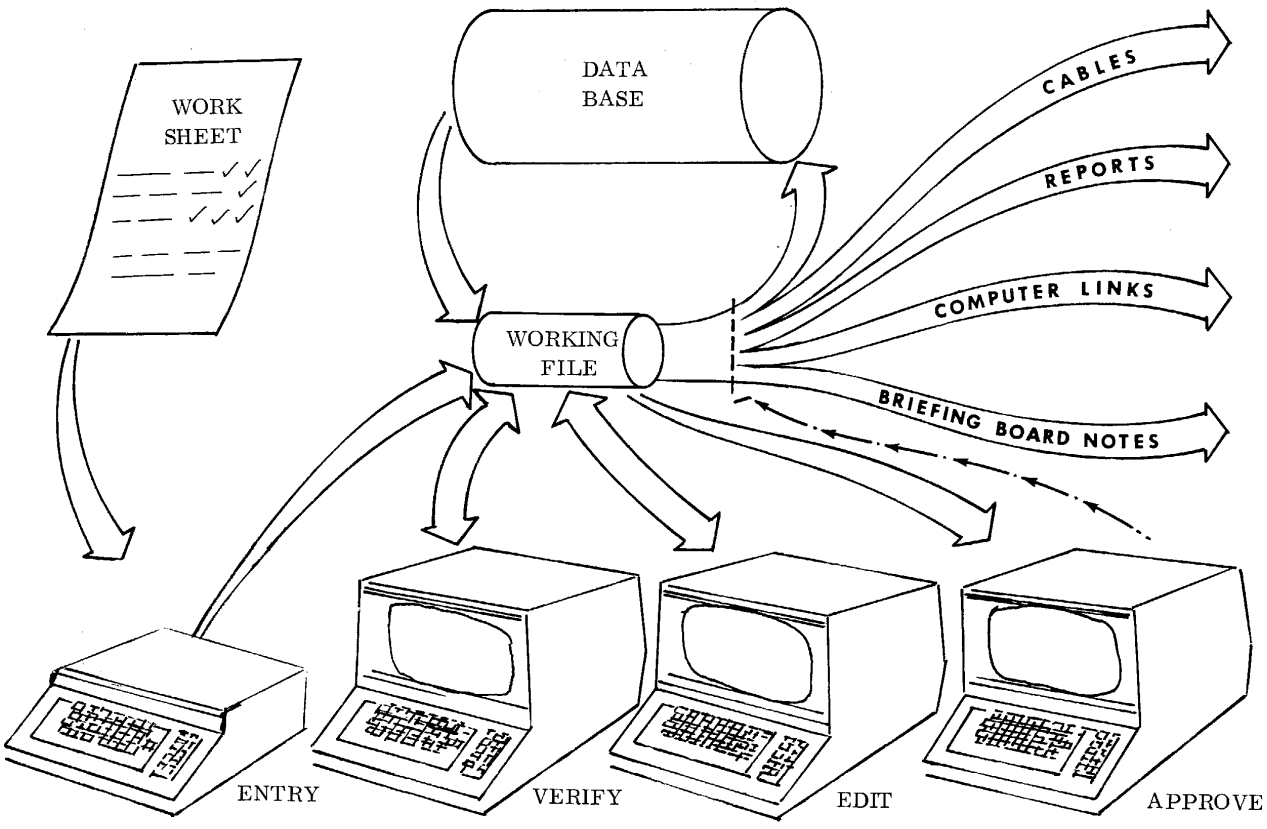


Figure 9 Output Cycle

TOP SECRET

25X1

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3
25X1

particular target readout has passed through that process. When the Mission Coordinator signifies his acceptance of the readout by depressing his special function key, the readout is used to update the data base and is available for output products.

The text for cables is prefixed with the necessary control information for the communication circuits and made available to the Communications Section immediately. The readout data for reports (cable and hard copy) are batched on a daily basis, sorted into publication order (sterilized and unsterilized versions), and used to prepare mats for printing and the output to communication circuits. If necessary, a proof copy of the report can be prepared prior to the preparation of the mats.

Provision is also made for providing selected portions of the readout data, such as OB data, available to other agencies in machine readable form for data base update purposes. This computer link can take the form of a communication circuit, in which case the data would be sent as soon as it was released by the Mission Coordinator; or it can take the form of magnetic tape or punched cards in which case it would be batched on a daily or mission basis.

The use of on-line consoles and random-access storage as part of the output cycle has two major advantages. First, since the information stays in the machine readable form throughout the cycle there is less paper passing, less elapsed time between operations, less reentry of data, and less chance for error. Second, since the output products are assembled from the random-access file, there is a great deal of flexibility in the content and organization of these products (changes are accomplished by making changes in the Report Format File used by the Report Generator Program). In addition, since the target header and other fixed data can be taken from the data base, there will be a significant savings in verification manpower.

4.4 Detailed Reporting

Third Phase readout is a retrospective, in depth study of an installation, group of installations or other topic. It differs from Immediate Readout in the time devoted to the readout and the time scale of interest. The IIS will contribute to Detailed Reporting in three areas, preparation, readout, and report preparation.

4.4.1 Preparation

When a requirement is received the files may be searched to identify all pertinent information and material, including work which may already have been done applicable to the requirement. After the requirement is validated and the extent of imagery interpretation necessary is defined, all the

25X1
Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

pertinent material, i.e., photography, reports, and other data is turned over to the PI. The identification of the pertinent material is accomplished by searching the data base, using the various indices that are available within the IIS. The computer search provides a significant advantage over manual search methods because the multiple access tags to information and the query capability will significantly decrease the probability of the PI overlooking information that should be included in the report.

4.4.2 Readout

This activity is illustrated on Figure 10. It is not significantly different from the illustration for immediate readout, the major difference being that the PI identifies material, by querying the data base and orders these materials for delivery to him. He also, of course, can query the data base to obtain supplementary information while he is studying the photography.

While not strictly a part of a readout process, Third Phase reporting often requires gathering, correlating, and analyzing a lot of data that has been developed on the basis of prior photo interpretation. The IIS concepts of data input and file structure of the National Data Base of Imagery Derived Information have been developed with an objective of improving the exploitation of this derived intelligence. Derived intelligence is defined, for these purposes, as information which can only be obtained through manipulation of the data concerning individual targets which is acquired by reading out one facility at a time. Examples of the manipulation of target data to derive additional intelligence are:

- Summarization of certain types of fixed facilities on a multi-target basis to identify National or organizational capabilities (e.g., summarization of the status and number of missile launch facilities in the USSR.)

- Summarization of the increases in certain articles of equipment or fixed facilities to confirm announced budgetary allotments for increased offensive and defensive capability (e.g., summaries of the new radar installations and associated equipment in a given area.)

- Data correlations to establish the changes in OB at various military installations which accompany changes at key installations during a military exercise (for example, changes in naval OB which accompany unusual concentrations of aircraft, to identify joint operations).

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

TOP SECRET

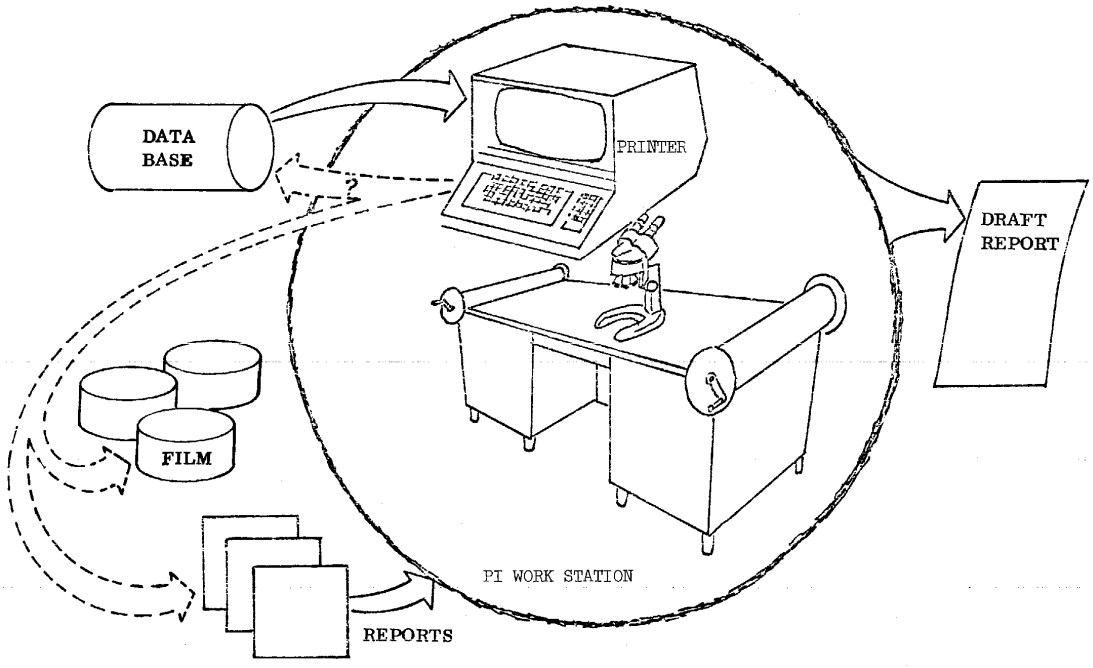


Figure 10 Detailed Reporting

25X1

TOP SECRET

25X1

SCR 352

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

These are only a few of the uses which summarization and correlation of readout data may serve. Some of this type of information is currently derived through manual summarization and/or correlation, however, the required investment in manpower discourages the development of these types of information in many areas. The provisions for selective access to certain data, in conjunction with the routine capability for complex statistical data handling characteristic of an automated information system, will significantly contribute to more effective utilization of the data base in detailed photo interpretation reporting.

4.4.3 Report Preparation

On-line data entry and editing capabilities are being provided to support Immediate Readout and can also be used for Detailed Reporting although there are differences in application. The major benefit to be achieved is the elimination of the retyping of successive drafts and the attendant proofreading required. This will be of particular benefit when updating Basic Intelligence Support Exploitation Reports and when extracting information from the data base for inclusion in reports. To support the former, the previous report is kept on magnetic tape so that the still pertinent portions can be incorporated in the new report. Computer programs are provided to support the latter.

4.5 National Data Base Activities

4.5.1 Data Base Inputs

The files of the National Data Base of Imagery Derived Information which NPIC is tasked with maintaining are the IDF, ODF, MPF, and EPDF. The inputs to these files come both from within the NPIC as a result of direct PI readout and from information generated by other agencies. In the paragraphs which follow, each file will be examined and the source of inputs discussed. Individual sectors of the IDF and ODF are examined in some detail, the structure of the MPF and EPDF does not require a detailed breakdown by sector.

a. Installations Data File

The Installations Data File contains information which is a result of image interpretation and information which can be used for the next readout of this installation. The updating of the image derived information is done immediately after approval of the readout data for each installation;

Header Sector - The information in this sector is assembled by CSD when a new target is entered into the system. The ultimate source of items such as COMIREX Number and BE Number is outside NPIC.

Brief Fixed Facility Description, Change Description, and Detailed Fixed Facility Description - The National Tasking Plan has stipulated that the agency which has Basic Intelligence Support Exploitation responsibility for an installation shall be responsible for its description of fixed facilities. Therefore the basic description will be the one contained in the latest third phase basic report. Descriptions for installations not covered by third phase reports will be updated as a result of first and second phase readout (See also Section 4.2.1).

Readout Requirements - These come from COMIREX. For non-COMIREX targets, the readout requirements are generated by NPIC.

Collateral Data - Information for this sector comes from CSD. As new or significant information becomes available for given installations, this sector is updated. Release of this information to the community may be restricted.

PI Notes - These are generated by NPIC PI's and are intended to direct the attention of the PI making the next readout of this installation. This sector is never sent out to the community.

Photo References - The inputs for this sector may come from two different sources. For film that is read out at NPIC these are a result of the readout activity. Photo references for film which was not read out by NPIC are taken from the photo interpretation reports received by the Center.

Source Data References - CSD supplies the information for this sector. It is updated as new references become known.

Concise Installation History and Complete Installation History - The inputs for these sectors are the PI readout of the installation. There will be an update each time that the installation is read out.

Formatted Fixed Facilities and Formatted Mobile Equipment - The

b. Objects Data File

The Objects Data File, as defined in the National Tasking Plan, contains information derived from reconnaissance imagery on weapons, weapon systems, associated equipment and objects of intelligence interest. The sectors of the ODF supply the information resulting from imagery interpretation along with identifying information and reference information about each object.

Header Sector - The information for this sector is supplied by CSD when a new, previously undefined, object is discovered.

Dimensional Data - This data is supplied by CSD. The basic source may be collateral, non-imagery derived intelligence, or Basic Support Exploitation done within or outside of the Center. If the Basic Support Exploitation was accomplished at NPIC, the dimensional data will be developed by TID. However, responsibility for updating the file rests with CSD.

Description - This sector's data is also supplied by CSD with the same basic sources as above.

Photo References - The data in this sector is a result of PI readout for film on which the object was read out at NPIC. This sector also contains references to film on which the object was seen, but which was not read out by NPIC.

Graphic References - These are references to drawings of the object and are supplied by TID for work which was done in-house at NPIC and by CSD for drawings originating outside the Center.

Readout Requirements - These will be established by COMIREX when a new object is discovered.

Source Data References - These will be supplied by CSD.

Location Information - This information is a result of PI readout of film. This provides a cross reference with the Installations Data File.

c. Mensuration Parameters File

The Mensuration Parameters File contains data which allows measurements to be made from the film. The inputs differ in form for each type of camera system. In all cases, the raw data is supplied by the collector. For KH-4 and INS sub-orbital missions the MPF data is generated from this raw ephemeris data while [redacted]

25X1D

25X1D [redacted] In the case of orbital
25X1D missions (KH-4 [redacted] the final attitude data is available only after data reduction of the stellar index camera system data. Until that time, nominal, or command, values are used.

d. Exploitation Products Data File

The inputs to this file result from products produced or received by NPIC which are a result of imagery exploitation. The maintenance of this file will be done by the same group which currently maintains the All-Source Index.

4.5.2 Query Capability

As was discussed in paragraph 4.5.1, the ability of the IIS to provide photo interpretation information to the community immediately after readout is one of the most salient features of the National Data Base activities. However, the data base input and maintenance features discussed in 4.5.1 are of little value in providing this immediate selectivity of information from the National Data Base unless the community user has some method of extracting data from the data base, and in turn has it presented to him at his work location, e.g. at DIA, State, CIA, or NSA. Extraction of data from the National Data Base is accomplished through the query capability of the IIS. The method of distribution of this extracted information to remote users is the COINS capability and will be discussed in paragraph 4.5.3.

The query capability of the IIS is perhaps the one capability of which the system user will be most conscious; many others, e.g. report generation, data entry, editing, approval, data base maintenance, etc. do not have the scope (number of users a capability interfaces with) and some IIS capabilities will never directly interface with the user at all, e.g. data base maintenance. Thus, the IIS query capability is of most interest to the user because it is his tool to communicate with the automated digital data base and extract the necessary information to accomplish his particular mission. This query capability is achieved in the IIS through the implementation of a query language known as PIC Information Retrieval Language (PIRL).

4.5.2.1 PIRL as a Query Language

There are three factors which must be considered in the design of a user oriented query language, the content and structure of the data base, the anticipated users, and flexibility to expand and accommodate change.

a. IIS Data Base

During the course of the design study, two tasks, reported in Tabs i

and j of Part III of this report were directed toward development of the data base. The first identified the contents of the IIS data base and the second ordered the contents into a structured entity. From these tasks, a set of terms has evolved which identify unique information elements such as COMIREX number and WAC/PIC number, etc. In effect, these terms establish the classes of search or retrieval parameters and relate directly to NPIC areas of interest. An initial list of these terms is shown in Figure 11. The terms, denoting the desired query response, provide the basis of the question that can be answered, hence the query language. The design is based on the premise that the terms should be derived initially from the contents of the data base and not by speculating on all possible queries.

b. Intended Users

The users of the IIS data base can be divided into two classes, those external to the Center and those in the Center. In the design of the IIS the users in the Center were further divided into two groups, a large number of "Special Users" who frequently require information from a limited portion of the data base, and a limited number of Information Specialists who are proficient in exploiting all aspects of the data base.

1. Special Users

Some groups within the Center will routinely want access to certain parts of the data base in carrying out their function. A primary example is the PI who, during readout, wants more information about an installation than appears on his Presentation. Provision is made in the query structure so that these groups can achieve a limited proficiency in exploiting the data base with a minimum amount of training and equipment. This is accomplished by having certain subsets of the basic query capability of the system available to them.

2. Information Specialists

Since it was not considered practical to keep all Center personnel, who require information from the data base, proficient in the use of the query language, Information Specialists are provided to aid them as required. The Information Specialists will be fully trained in the use of the system retrieval capability as it evolves over the life of the system, and hence will be able to serve the needs of Center personnel. It is anticipated that the information requests of Center personnel will be developed through a dialog with the Information Specialists as the Information Specialist queries the data base to see what is available.

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

TOP SECRET

CLASS	TERMS			
File	IDF (Installations)	ODF (Objects)	MTF (Moving Targets)	MPF (Mensuration Parameters)
Unit Record	Target Name COMIREX Number BE Number NPIC Number Country Military District Air Defense Zone Geo-Coordinates Category	Object Name US/NATO Code Numeric Code	Object Name	Mission Number
Sector	Header Brief Description Detailed Description Readout Requirements Collateral Notes Photo References Source Data References Fixed Facilities Mobile Readouts (OB) Brief History Complete History	Header Description Measurements Photo References Drawing Reference Source Data References Where Seen Essential Elements of Information	Description Measurement Photo References Where Seen	Pass Number

Figure 11 Query Language Terms

TOP SECRET

25X1

25X1

SCR 352

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

3. External Users

The external users considered are those who communicate with the IIS through COINS. Others, who may interface with the Information Specialists at the Center in person or by telephone, can be treated in the same manner as Center personnel.

c. Flexibility

Flexibility must be present in both the language itself and in the software which supports it. The software should be designed to allow for an open ended list of terms to permit the addition of new sectors or fields within files and the creation of new indices. In the language area, the primary purposes of initial implementation should be to provide convenient access to the data base and serve as an experimental tool for future query language development. As with any language, its use will uncover weak or deficient areas but the software should provide for their correction without major program modifications.

4.5.2.2 Query Language Usage

This section deals with the use made of PIRL by the three types of users identified in paragraph 4.5.2.1 and their method of interfacing with the system.

a. Internal User - Hardware Interface

The physical interface of the internal user with the IIS is the remote hardware configuration which allows him to query the system. This interface is achieved through two types of consoles:

1. The Standardized Usage System Interface Equipment Console (SUSIE console) whose operator is a trained information retrieval specialist with an intimate knowledge of the query language and data base plus a basic knowledge of the IIS system.

2. The PI Console whose operator is a PI with at least a general knowledge of the query language but little or no knowledge of the IIS system as a whole.

The use of a "function key" level of the query language makes is possible

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

facilitate time of entry for a query and allows an operator, like a PI, to query the IIS system without complete knowledge of the query language.)

b. Special Users

The most important, and perhaps the only, group of Special Users will be the PI's requesting additional information during Phase I and II readout. As described in Section 4.3.1, target related information will be available to the PI through target prediction and assignment procedures prior to mission readout. The goal is that this information be all that is required to read out approximately 75% of the covered targets. For the remaining targets, the additional information can be obtained using a simplified form of the Query Language.

An illustration of this simplified form or shorthand query, is as follows. A PI during immediate readout desires the past concise history of his target which is not included on the Presentation Sheet. This is achieved as follows:

1. The PI reads the five digit machine reference code from his Presentation Sheet and keys these five digits into the system at a console.

2. He then strikes the function key (e.g. upper/lower case key) which indicates the end of the machine reference number.

3. He keys in the letter "I" which requests the past concise history (see below for the list of Designator Keys).

With these six simple keyed entries (five numbers, a function, and a letter, the query is completed.

The complete set of function key requests for the IDF are as follows:

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

Machine Reference Number	Function Key	Designator Keys	Results from Query
NNNNN	f*	H	Installation Header
"	"	B	Brief Fixed Facility Description
"	"	D	Detailed Fixed Facility Description
"	"	E	Installation Readout Requirements
"	"	C	Collateral
"	"	N	Notes
"	"	P	Photo References/Coverage
"	"	S	Source References
"	"	O	OB History for Installation
"	"	I	Concise Installation History

* Any non-used key on a standard keyboard.

More complex queries (multi-term or multi-statement) are not covered by function key entries, but since the PI console has the same capability as the SUSIE console, it can accept such queries if the PI is sufficiently interested in learning the query language to generate them.

c. Information Specialist

Requests for more complex information than that described under special users are handled by an Information Specialist. There are two considerations that lead to this conclusion. First, it avoids the necessity that all PI's learn the IIS Query Language terms and structure initially, and, second, expertise will be built up by the Information Specialist as to data base content. These two considerations are to the PI's advantage when in search of additional information, because in contacting the Information Specialist a dialogue results which helps the PI define what information is needed and, at the same time, assists the Information Specialists in constructing the required terms of the query to obtain available information. Once a specific query has been established it is sent to the IIS via a keyboard device (i.e. SUSIE console), designating the requesting source. The IIS reacts by displaying the requested information on the original requestor's display and providing a printout if required. It is emphasized that the use of an Information Specialist will greatly reduce the number and frequency of improper, ambiguous and unanswerable requests to the data base and help identify information contained in, and available from, the data base.

d. COINS User

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

by the COINS user or may be generated by the special interrogation programs being developed as part of COINS I. The IIS output produced by the query statements may also be further processed by the COINS I interrogation programs.

4.5.2.3 Specific IIS Query Capabilities

The design task reported in Tab k of Part III established the query capabilities required in the operation of the IIS which were in turn submitted to the Center for approval. These approved requirements are:

1. The ability to retrieve sectors of a unit record by stating the file, unit record, and sector. This is accomplished by either:

   a. A term statement composed of terms that correspond to the file, unit record, and sector.

   b. A shorthand version of the above statement utilizing the "function key" technique. (See Section 4.5.2.2(b).

2. The ability to retrieve one or more fields of a sector by stating the file, unit record, sector, and field(s), e.g. the NAME and ELEVATION field in the header sector.

3. The ability to retrieve fields of a designated file, unit record, and sector on the basis of two qualifiers within the fields:

   a. Category type - in the case of mobile equipment.

   b. Time or Mission Number - can be designated also as a time span by the terms "after", "before", or "during" e.g. retrieve all photo references for a given target between March 1966 and June 1966.

4. The ability to establish the set of unit records containing designated terms in one or more classes such as the following:

   a. Country

   b. Military district

Note that in designating the term or terms in each class, provision shall be made for an "equal to" condition, using one or more terms, and a "greater than-less than" condition using two terms in an ordered list. In addition, when designating geographic location, provision shall be made for designating the size of a square centered at a point.

5. The ability to indicate to the originator of the query, the size of the set established by 4, before providing (and then only on request) a list of the set itself.

6. Provision of a dialogue capability, i.e. the ability to further refine the set established in 4. and 5. by adding or deleting to:

   a. expand the established set in order to retrieve on a broader scope

   b. restrict the established set allowing retrieval on a more limited scope.

The above elements represent the query capability provided by the IIS. Of particular importance is item 6 which allows a searcher to start with an ill-defined query and develop the definition of the data set desired on the basis of the information retrieved from the data base. A response time on the order of seconds is provided to make this capability a reality.

4.5.3 COINS Interface

In actuality, the interface of the COINS network to the IIS and therefore the Center is an extension to the query capability as discussed in Section 4.5.2. It allows NPIC users of COINS to direct queries to remote data bases such as NSA, CIA, DIA, and State and, in turn, allows personnel at these remote locations to direct queries to the IIS data base. Of prime importance to the Center is the capability to access, through COINS, remote data bases.

To accommodate an outgoing COINS query, the IIS will utilize the services of the Information Specialists (SUSIE operator) who are knowledgeable in the query language of the remote system e.g. a selected few of NPIC's SUSIE operators will be able to compose queries in the query language of NSA, DIA, CIA, etc., key them into the NPIC IIS system which, in turn, will direct these outgoing queries to the COINS switching computer

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

the three electronic/communications installations that are within this 1° square. Now, the PI wants to ascertain if NSA has any additional information of these three installations or indeed other known installations in this area. He then informs the SUSIE operator of the desired information of NSA. The SUSIE operator composes his new query in the TIPS language (NSA query language) and it is keyed in at the SUSIE console. Minutes later the NSA information is presented to the PI at his work station.

4.5.4 Data Base Maintenance

One of the major contributions supplied by the IIS is the currentness of the National Data Base of Imagery Derived Information. This currentness is achieved through an automatic on-line updating procedure (in the Data Management Package) which insures that once new photo interpretive data is approved in the reporting cycle, it is immediately entered into the National Data Base of Imagery Derived Information and properly indexed. In turn, this updated information is available to any internal remote console either internal to the Center (by using the query capability, see paragraph 4.5.2) or external to the Center (by use of the the COINS network). This automatic updating procedure is applied to the IDF, ODF, MTF, and EPDF files as a result of Phase I Reporting, Phase II Reporting and Detailed Reporting (Note the MPF is not updated in this automatic on-line fashion). All other updating to the IDF, ODF, MTF, MPF, and EPDF is included under the Data Management function as a manual input. This manual type of input includes corrections to all five files and adding of new material to the ODF, IDF, MTF and EPDF. Figures 12 and 13 show the automatic and manual sources of data for all files in the National Data Base.

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

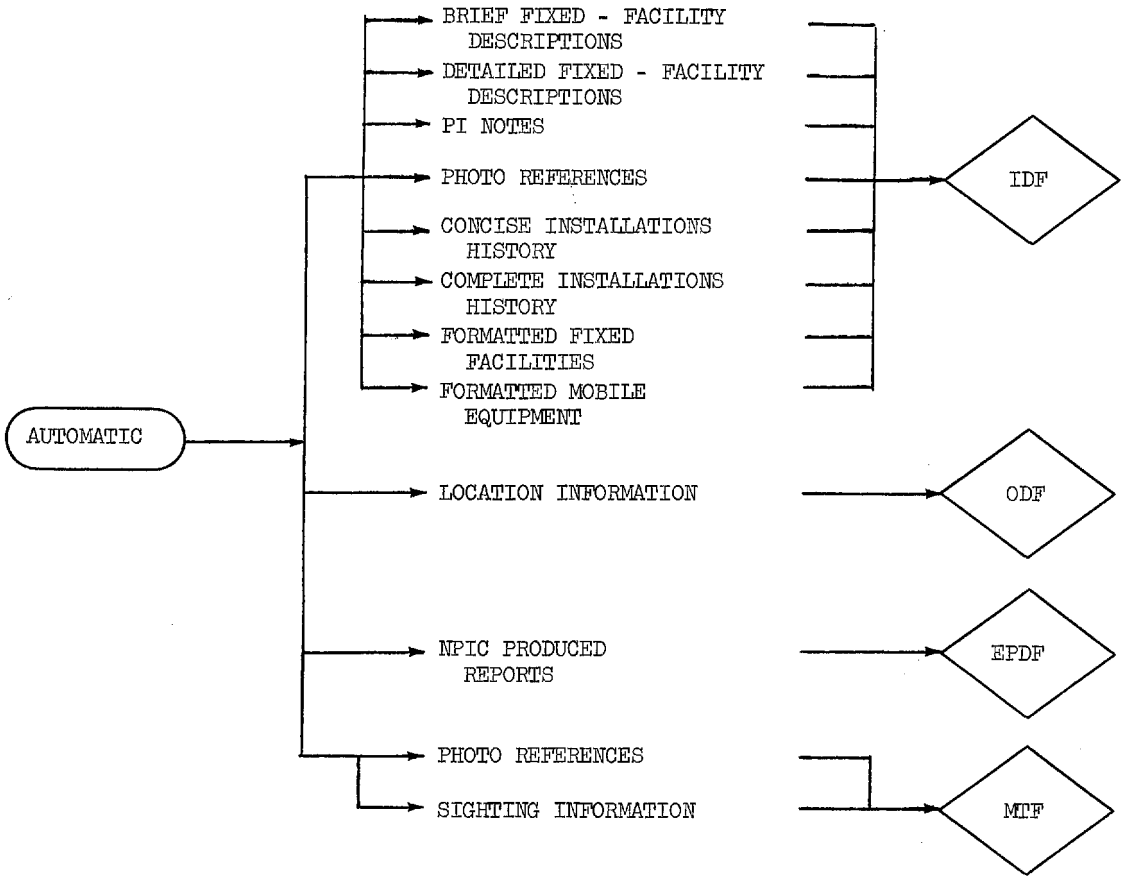
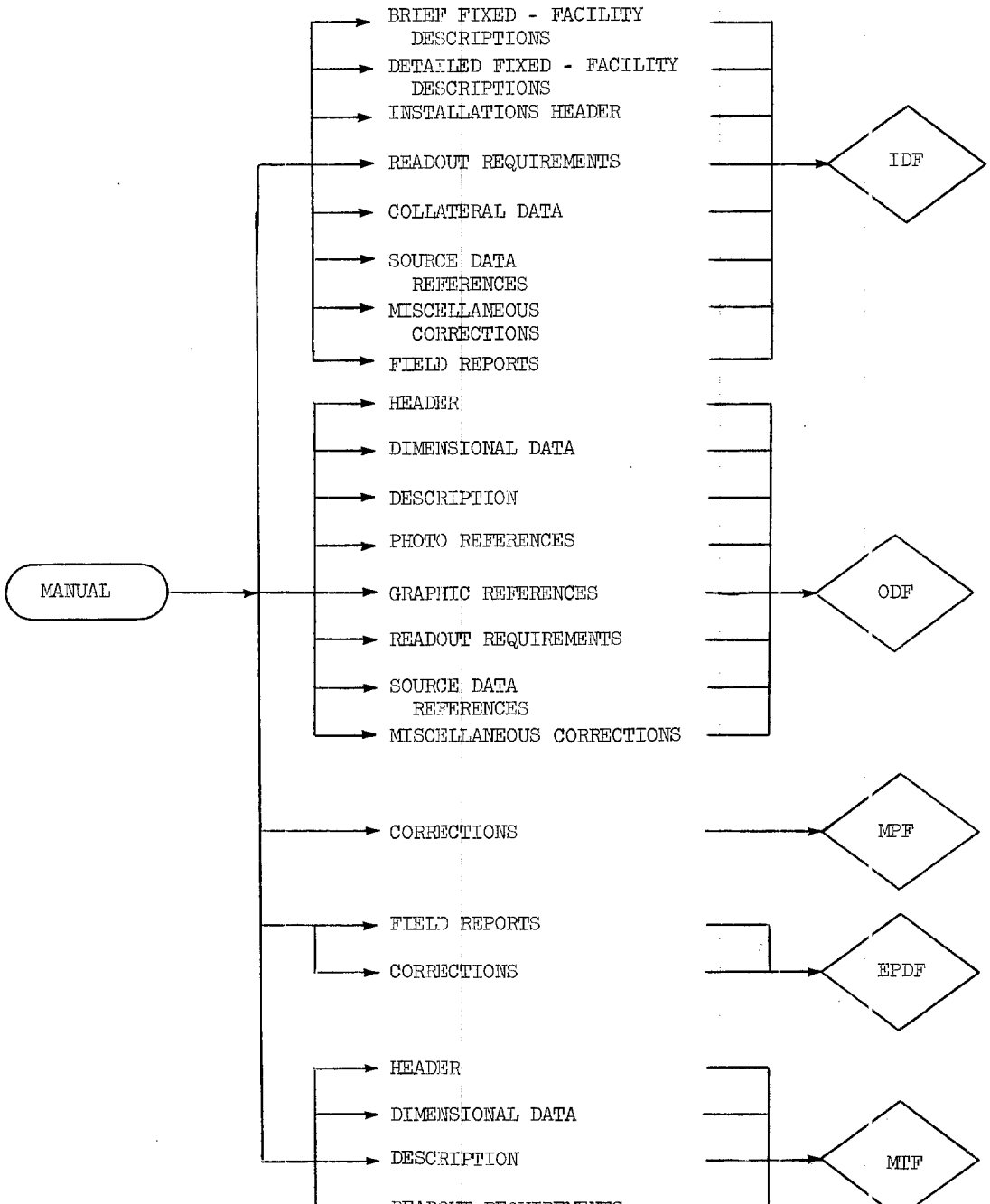


Figure 12 Automatic Sources of Data Base Updates

25X1 [redacted] TOP SECRET [redacted] 25X1
Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

SCR 352



5. DESIGN STUDIES, ANALYSIS AND ALTERNATIVES

This section presents the results of the analysis of the more important aspects of the IIS which were used in the selection of the major system alternatives. The basic Conceptual Design had previously been accepted by the Center and the effort expended in the detailed design was to ascertain that the system of the Conceptual Design was feasible at the detail level, that the system responses were within the desired bounds and that the costs were within reasonable limits.

It should be noted that the material in these sections records the results of the effort expended in the various studies and analyses; details of the procedures are contained in the Appendices A through G. Additional studies are described in Part III of this Report.

5.1 Data Base Considerations

The design of the IIS data base and its storage media has been a critical part of the design of the IIS itself. The determination of the content and organization of the files is discussed in Tabs i and j of Part III of this report. The determination of the storage media has been based on the three analyses described below, i.e., an estimation of the data base size, a determination of the data base utilization, and an investigation of the cost and performance of various combinations of storage media.

5.1.1 Size Estimates

5.1.1.1 Installations Data File

The starting point for estimates of the size of the IDF is the current Target Brief File (TBF). However consideration must be taken of the differences between the TBF and the IDF, the growth rate of the data about each installation, and the growth rate of the number of installations.

The current TBF consists of thirteen 2400 foot reels of magnetic tape (approximately 104 million characters) and covers about 21,000 installations. Of these, approximately 5,000 are COMIREX installations and the remainder are NPIC installations. Based on a ratio of three times as much data for the COMIREX installations as for the average NPIC, it can be shown that there are about 10,000 characters for the average COMIREX installation, and 3,300 for the average NPIC.

The estimates of size that have been made for the various sectors for each installation are shown in Table VII. The size estimates for many of the sectors are a function of the number of times the installations are read out, and the frequency of changes. The latter is also a function of

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

TABLE VII

IDF SIZE ESTIMATE

SECTOR	COMIREX INSTALLATIONS				NPIC INSTALLATIONS			
	Percent Having Sector	Characters/ Installation			Percent Having Sector	Characters/ Installation		
		Fixed	Initial	Annual Growth		Fixed	Initial	Annual Growth
Header	100	90	-	-	100	90	-	-
Brief Description	100	1,225	-	-	100	750	-	-
Detailed Description	20	12,000	-	-	1	6,000	-	-
Readout Requirements	100	200	-	-	100	200	-	-
Collateral Data	100	750	-	-	100	120	-	-
PI Notes	100	200	-	-	100	100	-	-
Photo References	100	-	1,500	784	100	-	800	387
Source Data References	100	300	-	-	100	150	-	-
Concise History	100	-	2,500	1,000	100	-	800	300
Complete History	100	-	10,000	3,000	100	-	3,300	1,000
Formatted Fixed Facilities	35	858	-	-	20	250	-	-
Formatted Mobile Equipment	25	-	8,000	4,000	10	-	2,000	1,000
AVERAGE TOTAL		5,465	16,000	5,784		1,520	5,100	1,787

TOP SECRET

25X1

TOP SECRET

25X1

SCR 352

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

the detail of readout. The Input-Output Loading Factors (Section 3) indicate the average numbers of missions/year, the average numbers of installations/mission and the total numbers of installations. The summations of these indicate that there will be about 45,600 COMIREX installation readouts per year and 112,800 NPIC's. Therefore the average COMIREX installation is read out eight times per year and the average NPIC four. If the percentage of time that a significant change is noted is taken as about 25% for COMIREX's and 15% for NPIC's the average number of changes per year per installation will be about 2.0 and 0.6 respectively. When the readouts by other agencies that are entered into the IDF are considered, these numbers can be increased to 2.5 and 1.0 respectively.

The estimate of the Brief Description is based on a study of current installation descriptions and an allowance for statements of change that are appended to the basic description. Since all of the COMIREX's are covered by Basic Intelligence Support Exploitation Reports that are updated on the average, once every year, there will be an average of 1.25 change statements per COMIREX installation. The number for NPIC installations has been taken as 0.5. These change statements are estimated at 500 characters each for COMIREX's and 300 for NPIC's. The basic descriptions are estimated at 600 characters for both types.

The need for Detailed Descriptions and the size estimates for it, Readout Requirements, Collateral Data, PI Notes, and Source Data References are all based on a study of records in the current TBF. Although these are updated, they are all fixed in average length because the information is replaced rather than accumulated.

Photo references, on the other hand, are cumulative and the amount of data is proportional to the amount of coverage. The estimates are based on 80 characters per reference, the number of readouts per year, and 1.1 photo references per readout.

The Concise History contains the initial readout and an accumulation of all the change statements. It has been estimated that the concise histories for those installations in the current TBF would be about one quarter the size of the total current TBF. The Complete History is also cumulative and represents almost the total volume of the current TBF.

Formatted Fixed Facilities consists of one set of 130 characters for each type of facility (e.g., launch sites or barracks) that is updated as changes are noted rather than accumulated. The estimates given represent

25X1

TOP SECRET
Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

of 50 characters per object sighting.

The totals indicate that each COMIREX and NPIC installation in the current TBF will start with 21,465 and 6,620 characters and grow at rates of 5,784 and 1,787 characters per year, respectively. In addition, as each new installation is added, it will increase the file size by 5,465 and 1,520 characters and the growth rate by 5,784 and 1,787 characters per year, respectively.

The target growth rate can be determined from the current numbers and the numbers given in Section 3 for the early 70's to be 250 per year for COMIREX and 3,500 per year for NPIC's. (Since the COMIREX list has deletions as well as additions, the figure of 250 per year should be taken as the net growth.) With these rates and the totals on Table VII, the initial size of the IDF will be 213 million characters with an initial growth rate of 64.2 million characters/year. Further, as the numbers of installations increase, the growth rate will increase by 7.69 million characters/year each year.

It will be noted that the initial size is bigger than the current TBF because the current TBF is included as the Complete History and the IDF contains installation descriptions which are not in the TBF as such. It is not anticipated however that the Complete History will be kept in on-line storage.

5.1.1.2 Mobile Target File

Since this or an analogous file is not presently maintained there is no objective base from which to extrapolate its probable size. However, an estimate of the maximum size can be based on the following:

a. Each object can be identified and described in less than 1000 characters.

b. Each object will be seen on less than 20% of the buckets of film.

c. There will be less than 52 buckets of film per year.

d. Each sighting can be described in less than 100 characters.

e. The information is of interest for

On this basis the file will be less than four million characters. Since this number is small with respect to the remainder of the data base, it is used as the estimated size.

5.1.1.3 Objects Data File

The exact content of the ODF has not as yet been defined. However, the following estimates of the size of the various sectors for a typical object appear reasonable:

Header	120	characters
Dimensional data	30	
Object description	300	
Photo references	60	
Graphic references	20	
Readout requirements	200	
Source data references	70	
Location information	1000	
Total	1800	characters

The Input-Output Loading Factors indicate the potential for 1000 object classes. It has been assumed that the average object class will have about five objects and that 20% of the total would be implemented initially and 20% in each succeeding year. On this basis the initial size of the ODF will be 1.8 million characters and the growth rate will be 1.8 million characters per year.

5.1.1.4 Mensuration Parameters File

The size of the MPF is a function of the number and type of missions flown. These are given in Section 3. Of the aircraft missions it was assumed that 60 per year were INS equipped and would have data in the file. The number of characters per mission were taken as:

Search	1,200,000

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

Taking an initial file of the equivalent of 36 search and 20 spotting missions, the initial size of the MPF will be 48 million characters and the growth rate will be 27.6 million characters per year.

5.1.1.5 Exploitation Products Data File

The size of the sectors in each record in the EPDF has been estimated to be:

Header	80 characters
Location data	54
Alternate titles	150
Total	284 characters

There are currently 80,000 reports in the All-Source Index and this is growing at the rate of 11,000 per year. Therefore the initial size of the EPDF will be 23 million characters and the growth rate will be 3.1 million characters per year.

5.1.1.6 Total Size

The total size of the IIS data base is the sum of the five files discussed above. The initial size is 290 million characters and the size as a function of time is shown in Figure 14. This indicates that the size of the data base will increase at a quadratic rate. The quadratic rate is caused by the linear increase in the number of installation records with time and the fact that the size of these records also increases linearly with time. Also indicated on Figure 14 is the contribution of each of the files to the total size.

A significant portion of the total is the Complete History data in the IDF. Another significant portion is data that is over two years old (i.e. parts of the Concise History, Photo References, Formatted Mobile Equipment, Mensuration Parameters and Exploitation Products data). The relation of these two portions to the total is shown on Figure 15. This indicates that if the Complete Installation Histories are kept on hard copy or off-line magnetic tape and if a certain set of the perishable data is purged from the active storage (e.g. converted to hard copy or magnetic tape), the active portion of the data base grows at a rate of 22

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

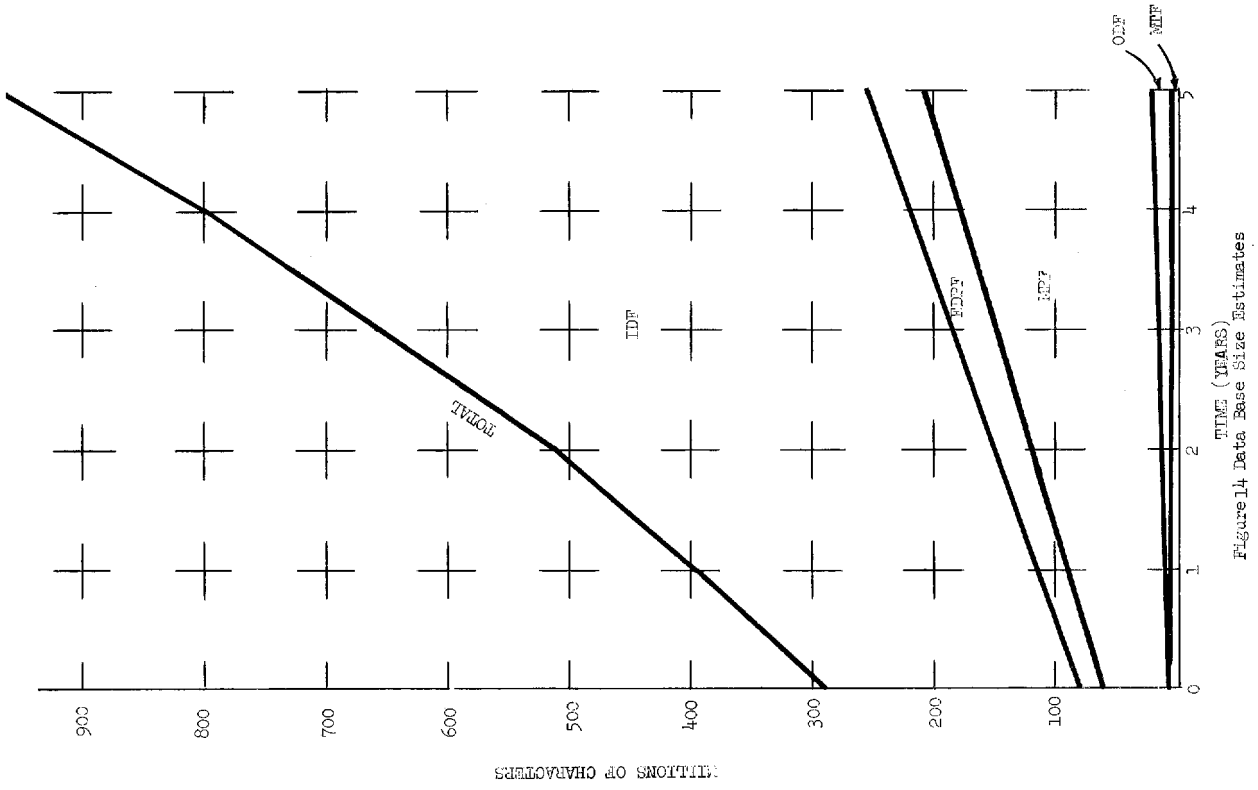


Figure 14 Data Base Size Estimates

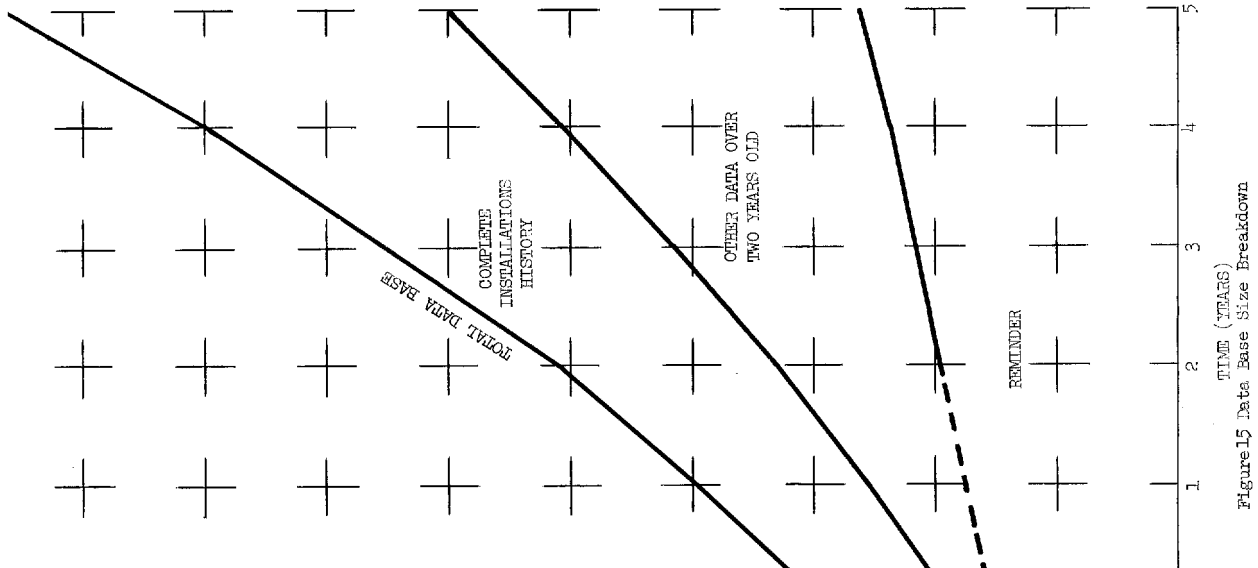


Figure 15 Data Base Size Breakdown

25X1 [redacted] TOP SECRET [redacted] 25X1
Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

5.1.2 Utilization

A necessary step in designing the data base is an investigation of the usage that the various parts of the data base will experience. Such an investigation would indicate what portion of the data base receives the most use and hence should be considered for the most accessible portion of the storage. Conversely, it should also indicate what portion of the data base is little used and could be put in less accessible storage without seriously affecting system performance.

Since a data base of the form provided by the IIS does not exist there is no actual data describing its use. Therefore the various operational modes of the Center were studied to determine how they could be supported by each sector of each file of the data base. An attempt was also made to indicate how the use of the perishable data (e.g. photo references) would vary with age. The operational modes studied were first phase reporting, second phase reporting, third phase reporting, and general queries such as those to be received over COINS. The request rates were established for each of the sectors under each of the operational modes. The workload in each of the operational modes was established from the Input-Output Loading Factors to be:

First phase	-	3800 target readouts/month
Second phase	-	9400 target readouts/month
Third phase	-	600 reports/month
General queries	-	20,000/month

The analysis is described in Appendix B along with the input data and the detailed results. A summary of the results is given on Table VIII and Figure 16. Table VIII lists the various sectors of the data base in order of use and gives the cumulative fraction of the total number of requests for all sectors up to that place in the list. Figure 16 shows the fraction of the total number of requests that can be answered by the corresponding fraction of the data base. This curve indicates that there is a small fraction of the data base that can answer a large portion of the requests. For instance, 25% of the data base can answer 97% of the requests. Note also that the least used 40% of the data base is needed for only 0.1% of the requests.

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

did vary, the curve on Figure 16 did not change substantially. Therefore it may be concluded that in actual practice, the order of use may vary from that shown on Table VIII but the shape of the curve on Figure 16 will be essentially unchanged.

TABLE VIII

USE OF IIS DATA BASE

	Sector	Cumulative Fraction
1.	ODF High-Dimensional Data	.0124
2.	ODF High-Header	.0501
3.	ODF High-Photo References	.0594
4.	IDF COMIREX Header	.1298
5.	ODF High Source Data References	.1349
6.	ODF High-Graphic Reference	.1362
7.	ODF Low-Photo References	.1555
8.	ODF Low-Dimensional Data	.1628
9.	ODF High-Readout Requirements	.1692
10.	ODF Low-Header	.1938
11.	ODF High-Object Description	.2004
12.	ODF High-Location Information	.2222
13.	EDPF-Location Data-Under 2 Years	.2446
14.	IDF NPIC Header	.3131
15.	IDF-COMIREX-Formatted Fixed Facilities	.3611
16.	EDPF-Alternative Titles-Under 2 Years	.4055
17.	IDF-NPIC-Formatted Fixed Facilities	.4279
18.	EDPF-Header-Under 2 Years	.4422
19.	IDF COMIREX-Readout Requirements	.4558
20.	ODF-Low-Source Data References	.4590
21.	ODF Low-Graphic References	.4598
22.	IDF-Inactive-Header	.4652
23.	IDF-NPIC-Source Data References	.4959
24.	IDF-COMIREX-Source Data References	.5080
25.	ODF-Low-Location Information	.5333
26.	IDF-COMIREX-PI Notes	.5400
27.	IDF COMIREX-Brief Description	.5811
28.	IDF-NPIC Readout Requirements	.6114
29.	IDF-NPIC-PI Notes	.6251
30.	IDF-COMIREX-Photo References-Under 2 Years	.6661

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

TABLE VIII (cont'd.)

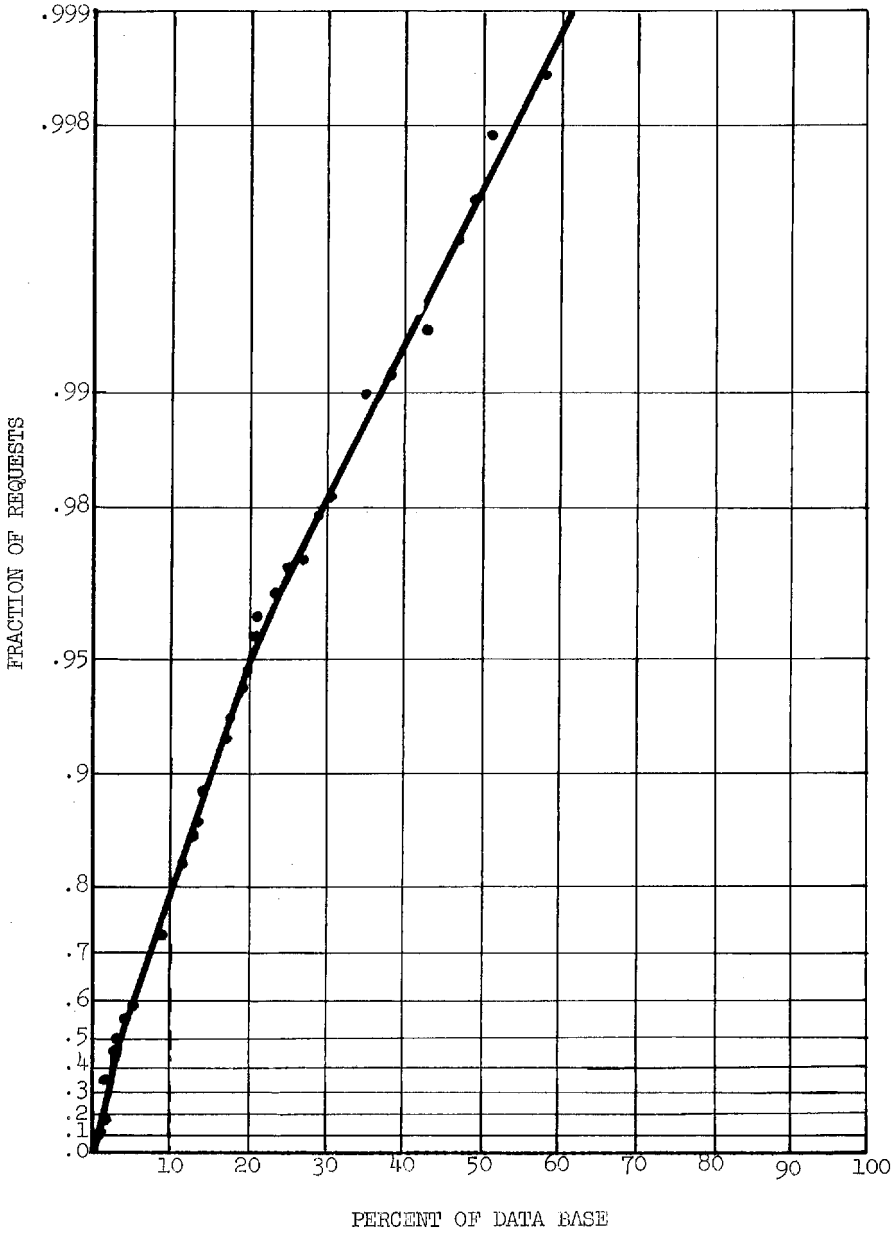
36.	IDF-NPIC-Collateral	.8578
37.	IDF-NPIC-Formatted Mobile Equipment-Under 4 Years	.8817
38.	IDF-NPIC-Formatted Mobile Equipment-Over 4 Years	.8897
39.	IDF-COMIREX-Formatted Mobile Equipment-Under 4 Years	.9200
40.	IDF-COMIREX-Collateral	.9255
41.	IDF-COMIREX Concise History-Under 2 Years	.9392
42.	IDF-NPIC-Detailed Description	.9412
43.	IDF-COMIREX-Formatted Mobile Equipment-Over 4 Years	.9508
44.	MPF-0-3 Months	.9582
45.	IDF-COMIREX-Detailed Description	.9644
46.	IDF-NPIC Concise History-Under 2 Years	.9708
47.	IDF-COMIREX-Photo References-Over 2 Years	.9762
48.	EDPF-Alternate Titles-Over 2 Years	.9795
49.	IDF-NPIC-Photo References-Over 2 Years	.9869
50.	EDPF-Header-Over 2 Years	.9882
51.	IDF-COMIREX-Concise History-Over 2 Years	.9903
52.	IDF-NPIC-Concise History-Over 2 Years	.9933
53.	IDF-COMIREX-Complete History-Under 2 Years	.9960
54.	IDF-Inactive-Remainder	.9968
55.	MPF-4-12 Months	.9978
56.	IDF-NPIC-Complete History-Under 2 Years	.9985
57.	IDF-COMIREX-Complete History-Over 2 Years	.9992
58.	IDF-NPIC-Complete History-Over 2 Years	.9999
59.	MPF-1-2 Years	.9999
60.	MPF-Over 2 Years	1.0000

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

SCR 352



5.1.3 Storage Media

Sections 5.1.1 and 5.1.2 established the approximate size of the data base and the relative use of its various parts respectively. This provides the background to investigate the various combinations of storage media that might be used to contain the data base and to evaluate the relative performance of each.

5.1.3.1 Selection Criteria

The criteria for comparisons among the various storage configurations has been the cost and average response time. The reliability of the components in the configuration has also been considered in that it affects the average response time and cost.

Average response time has been taken as the measure of effectiveness since, for a large fraction of the data requests, the requestor is a person for whom time is important. As the response time is reduced from minutes down through seconds, the person becomes better and better able to carry out a dialogue with the data base to arrive at the desired information without losing his train of thought. However, it is felt that there is little gain in effectiveness in reducing the response time below a few seconds.

The other important consideration is reliability. Since the data base is used in an operational environment that functions continuously, the data base must also operate essentially continually. The measure of "essentially continually" has been taken as at least 99.9% of the time. This means that the total data base or any part of it can only be inoperative less than 2 hours per year of single shift operation. Since the usual mass storage devices, i.e. tape frames and drums, discs, do not approach this performance singly, much less in combinations, some form of backup or redundancy must be provided. Thus reliability can be considered by including the costs of the backup equipment with the costs of the primary equipment and including the response time of the backup mode, weighted by its probability of occurrence, in the average response time. In costing backup equipment it was assumed that the total computer system would have one magnetic tape control unit and ten magnetic tape units that could be used provided they are not required more than 5% of the time.

5.1.3.2 Types of Storage

purge the active file of the less useful data. Although the purged data is printed out, reference to its location (i.e. volume and page) is retained in the IIS indices. If the hard copy is accessible by the users, either directly or via a telephone, the response time for the hard copy data should be less than one-half hour. This can be accomplished even for
25X1A users outside [redacted] by distributing copies of the data when it is printed out.

Next lowest in cost is the off-line tape. In this case the purged data is written on tape and the reel put in a tape library. The access time is about the same as with hard copy but the data is always available in machine readable form.

Even on-line magnetic tape is an inherently slow form of storage since the data can only be accessed serially. It can be shown that there is not sufficient capacity with magnetic tape to handle the IIS requests on a first-in, first-out basis. Therefore requests must be batched and the data base scanned to retrieve the data for all the requests being held in the batch. If indices to the data are available, the time to scan a data base contained on several tapes can be reduced by providing several tape control units.

Drums provide fast access since all the data is available on each revolution of the drum. The speed advantage of drums is lost, however, if a very large number of records must be searched to find the desired data. Therefore a necessary part of the drum storage system are indices to the data.

5.1.3.3 Storage Comparisons

The characteristics of the various devices that were used in this analysis are summarized below.

	Cost	Capacity	Availability*
25X1A Magnetic Tape Unit	[redacted]	17×10^6 characters/reel	95%
Tape Control Unit		16 tape units	98%
Fastrand Drum		125×10^6 characters	92%
Drum Control Unit		8 drums	98%

The primary concern with respect to the tape units is to achieve a small response time. As indicated in Appendix C, the response time is

25X1 [redacted] TOP SECRET [redacted] 25X1
Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

computer configuration.

The primary concern with respect to using drums is to achieve the required reliability. The ten tape units in the computer configuration can be used as a backup but this involves a significant penalty in response time. Therefore the important design parameter was the number of drums to be provided in excess of those required to store the data assigned to drums.

The objective of the analyses was to determine the average storage system response time and hardware costs for several configurations of on-line magnetic tapes and large drums. The response time is the average response time for all queries of the data base whether the necessary data be stored on-line or off-line. The average response time also includes an allowance for operation in the appropriate failure modes. Hardware costs include backup equipment as well as the prime equipment.

The details of the analyses are given in Appendices C and D and results are presented on Figures 17 and 18 for tape systems and drum systems respectively. In both analyses one of design parameters was the fraction, f, of the total data base that was assigned to on-line storage. This fraction was used in conjunction with the function plotted on Figure 16 to determine the fractions of the requests directed towards the on-line and off-line storage. These fractions provide the means for determining an average storage system response time, weighted in accordance with the relative use of the two kinds of storage.

The other design parameter in the tape system study was n, the number of tapes per tape control unit. Figure 17 shows the decrease in response time as n is decreased and the corresponding increase in system cost if f is held constant. The cost increase corresponds to the cost of the tape control units.

A line representing a fixed product of cost and time has been drawn on Figure 17. This represents an even trade-off between time and cost. It indicates that the tape storage system with the minimum cost-time product is one with one tape per control unit and one-third of the data base on-line. Such a storage system would have an average response time of about
25X1A 64 seconds and a cost of [redacted] (Note: This response time may be low by a factor of 2 or 3 because an assumption made in Appendix C on timing may not be valid with one tape per control unit.)

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

TOP SECRET

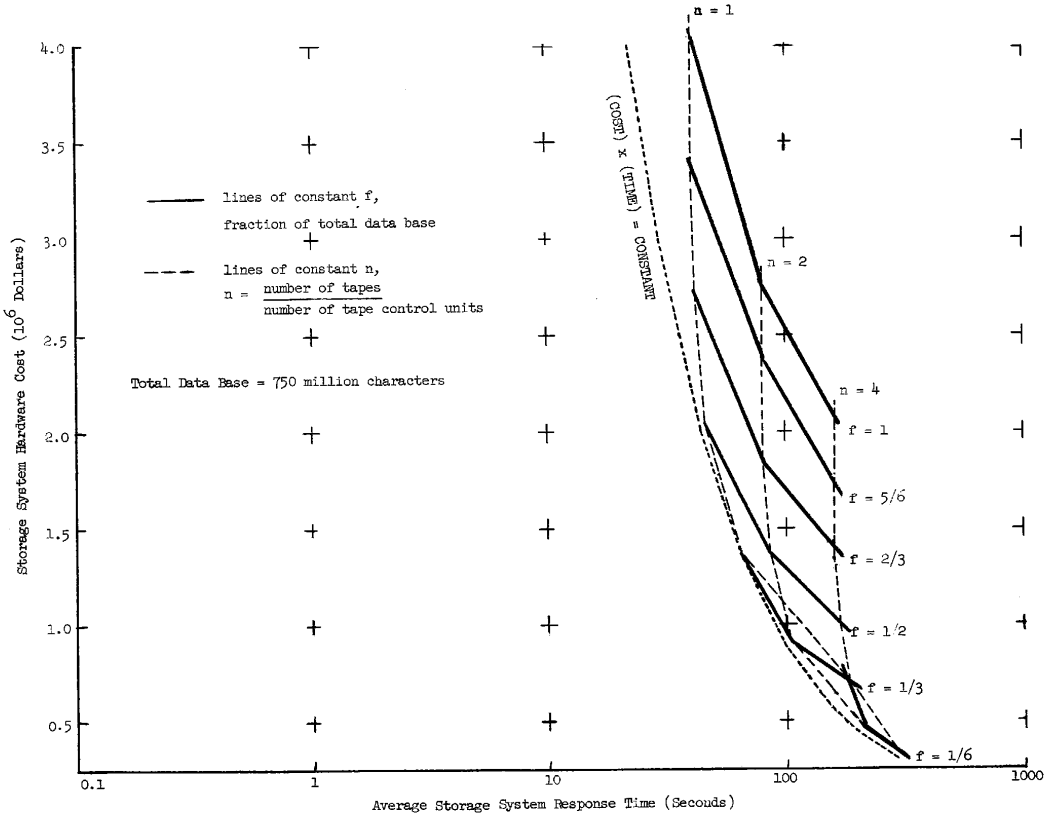


Figure 17 Characteristics of Tape Storage

TOP SECRET

25X1

25X1

SCR 352

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

TOP SECRET

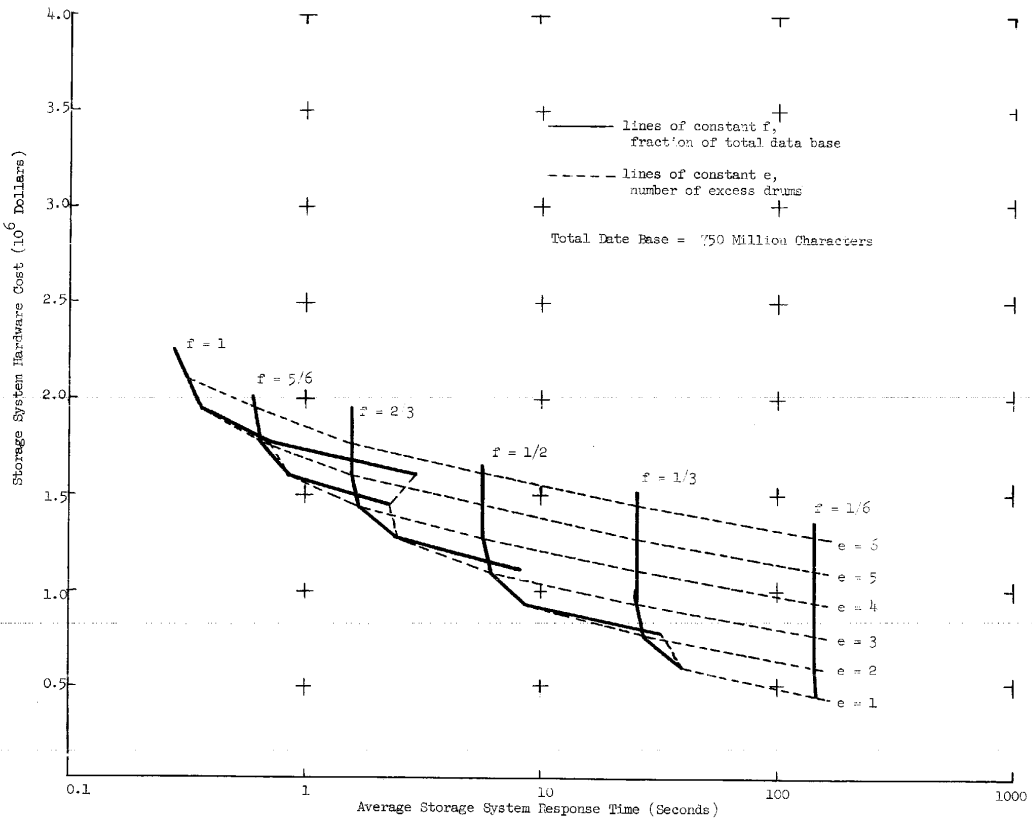


Figure 18 Characteristics of Drum Storage

25X1

25X1

TOP SECRET

SCR 352

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

provided from the remainder of the computer system. Figure 18 indicates that there is not much advantage in this unless a large portion of the data base is assigned to on-line storage. There are two reasons for this. First, with only a small portion of the data base on-line, the number of drums is small and the probability of a drum failure is correspondingly small. Second, and more important, with small on-line storage, the bulk of the average response time is contributed by the time to retrieve data from off-line storage and is not affected by improvements in the on-line storage response time.

The configuration with the smallest time-cost product is all data on-line with six excess drums. This configuration has a response time of 0.31 seconds. However it is realized that there is not an even trade-off between time and cost in this time range. There is no universal answer to the question of establishing a response time below which there is no further benefit. However considering human reaction time and the operations of a person at a query console a limit of two to three seconds seems reasonable although some people will argue for less. Such a limit can be incorporated in a cost/effectiveness parameter by using the product of time plus five seconds and cost. This has the effect of saying regardless of how much the response time is reduced below five seconds, the effectiveness factor can at most increase by a factor of two. Figure 19 gives this cost/effectiveness parameter for each of the tape and drum configurations.

As indicated by Figure 19 the cost/effectiveness is a function of both the design configurations and the amount of data on-line. These plots indicate that for tape systems the "best" mode of operation is with a relatively small amount of the data on-line. (The "best" point is the same one identified on Figure 17.) However for drums the reverse is true. Although there is little difference between the configurations with two-thirds and five-sixths of the data on line, the former has a slight advantage. This system has the following characteristics:

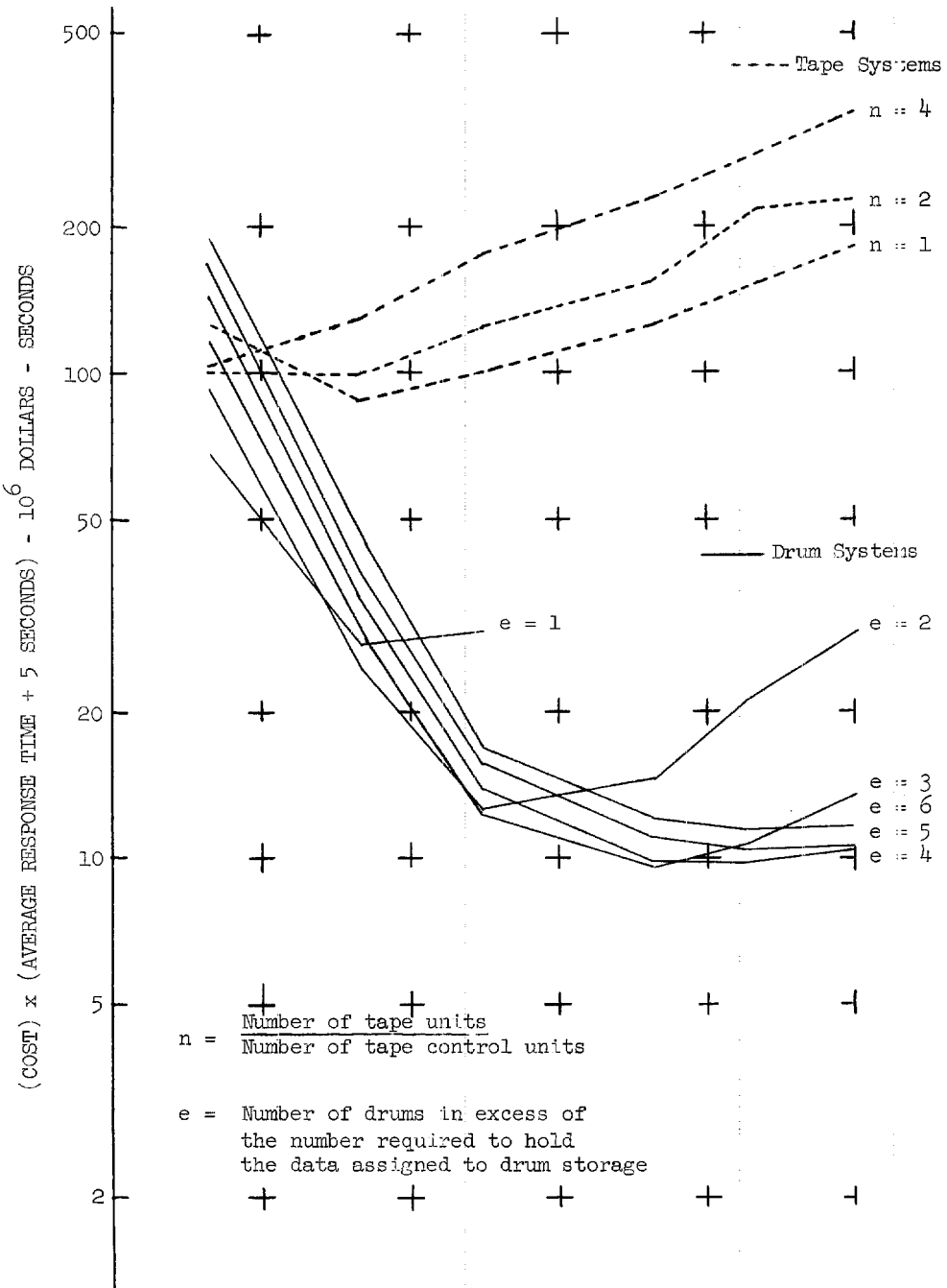
Cost	[redacted]
Average Response Time. . . .	2.5 seconds
Amount of data on-line . . .	67%
Percent of queries answered on line	99.85%
Number of drums used	4
Number of excess drums provided	3

25X1A

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1



No formal analysis of the best form of off-line storage has been made. However, it would seem appropriate to maintain the MPF data on magnetic tape in a tape library so that it is available in machine readable form. The histories on the other hand might better be printed and distributed to libraries where they could be available for study.

Some consideration was also given to a three-way combination wherein drums would handle the queries directed to the most used portion of the data base,the least used portions would be in off-line storage, and on-line tapes would be used in the middle range. However since the cost per unit of storage is about the same for drums and tapes, and since the effective incremental cost for drums decreases as the data stored on drums increase, there is no advantage in such a configuration.

5.2 PI Presentation Sheet Content

General familiarity with the readout process and the responses to a questionnaire given to the PI's (see Part III, Tab b) indicate that less than one-third of the total data on a given target in the IDF is used during the average readout. Therefore, there might be advantages to giving the PI less than the total amount of data prior to the readout. However it is considered impossible to predict exactly which data elements will be required on each target. Therefore, if less than the total data is given, provision must also be made for the PI to request additional data that might be required. However, the saving in PI time should compensate for the cost of providing the request capability. An analysis was conducted to determine the effect such a practice would have on the time a PI spends assimilating this data and to determine the number of requests for additional data that would result.

The approach was to divide the total data into sets in accordance with its use as shown on Tables IX and X. The probability, or relative frequency, of use of each set was established on the basis of responses to the PI questionnaire and other familiarity with the readout process. It was assumed that, as a minimum, the header information for all predicted targets would be given to the PI's. The amount of time the PI devotes to the data was assumed to be proportional to the amount of data routinely distributed to him plus the average amount of data requested. An allowance was also made for the time required to make requests and obtain responses.

The mathematical analysis used is given in Appendix A and the results

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

TABLE IX

DATA UTILIZATION - FIRST PHASE READOUT

DATA SET	CONTENT		AVERAGE SIZE (Characters)	PERCENT AVAILABLE	PROBABILITY OF USE IF READOUT
0	Header		90	100	all predicted
1	Readout Requirements Collateral Data PI Notes Mobile Equipment - last 2 missions	200 750 200 250	1400	100	1.00
2	Recent Photo References - last 5 entries		400	100	.75
3	Basic Description Changes Formatted Fixed Facilities	600 625 300	1525	100	.75
4	Source Data References		300	100	.30
5	Concise History		6000 *	100	.10
6	Complete Photo References		4200 *	100	.05
7	Detailed Description		12000	20	.04
8	Complete Mobile Equipment		22000 *	25	.01

* Extrapolation to 3 1/2 years from the present.

TOP SECRET

TOP SECRET

25X1

25X1

SCR 352

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

TABLE X

DATA UTILIZATION - SECOND PHASE READOUT

DATA SET	CONTENT		AVERAGE SIZE (Characters)	PERCENT AVAILABLE	PROBABILITY OF USE IF READOUT
0	Header		90	100	all predicted
1	Readout Requirements	200			
	Collateral Data	120			
	PI Notes	100	470	100	1.00
	Mobile Equipment - last 2 missions	50			
2	Basic Description	600			
	Changes	150			
	Formatted Fixed Facilities	50	800	100	.95
3	Source Data References		150	100	.30
4	Recent Photo References - last 5 entries		400	100	.30
5	Concise History		1850 *	100	.05
6	Complete Photo References		2150 *	100	.05
7	Complete Mobile Equipment		5500 *	10	.05
8	Detailed Description		6600	1	.002

* Extrapolation to 3 1/2 years from the present.

TOP SECRET

TOP SECRET

25X1

25X1

SCR 352

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

TOP SECRET

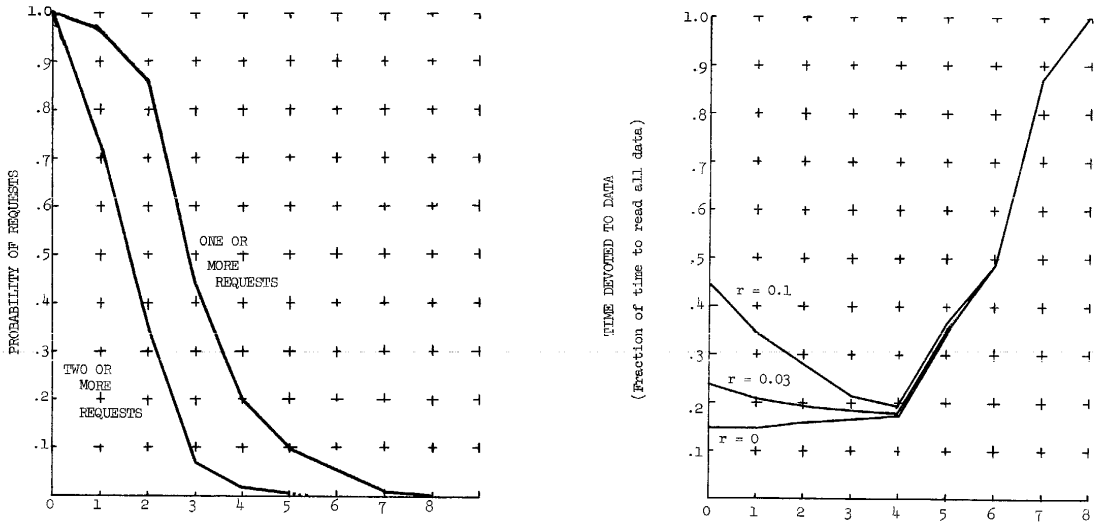

PROBABILITY OF REQUESTS
ONE OR MORE REQUESTS
TWO OR MORE REQUESTS
Number of Sets Given Routinely
TIME DEVOTED TO DATA
(Fraction of time to read all data)
r = 0.1
r = 0.03
r = 0
r = time for a request / time to read all data

Figure 20 Effect of Varying Presentation Content

25X1

25X1

SCR 352

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

read out. The right set of curves represents the amount of time the PI would spend with the data for various amounts of time to obtain a response to a request. Time has been normalized using the time to read all the data as a base. The probabilities of use given on Table IX indicated that on the average he requires approximately 15% of the total data to complete the readout. Thus, as more data is given routinely, time increases because data is being given that is not required.

The results indicate that approximately 80% of the time that would otherwise be devoted to data can be saved by routinely giving the PI only the header and the four sets of data most likely to be used. If the base time is 15 minutes, the saving is 12.5 minutes per target. The results also indicate that, if these four sets are given, additional data will only be required for 20% of the targets read out. Therefore, it was decided to routinely give the PI the header data and data sets 1 through 4 on Table IX and the capability to request additional data. The number four was chosen over some lesser number to minimize the number of requests without significantly increasing the PI time. The characteristics of this mode of operation are summarized on Table XI.

The analysis indicates that when four sets are given the request rate will be very low, less than one per hour per PI, and that the PI time is fairly insensitive to variations in the time to request additional data. This means that the request device need not be immediately adjacent to the PI work station; he could well afford to walk across the room to use a device shared by a group of PI's. (The load created by six or eight PI's is not sufficient to cause an appreciable queue.) Since the data sets to be requested are the longer ones (several thousand characters each), printers should be considered as the device to present the request response to the PI. This will produce hard copy that the PI can carry away and study rather than the character-limited, volatile display of a CRT. The speed of the printer should be somewhat faster than the reading speed of the PI so that he is not waiting for data. A Kleinschmidt or equivalent printer should be sufficient. Such a device would be shared among six or eight PI's doing first and second phase readout and would cost about [redacted] as described in Tab c in Part III. Allowing for as many as 150 PI's doing first and second phase readout and as few as 6 PI's per request printer, 25 would be required for a total of [redacted]

25X1A 25X1A

To determine whether or not such a cost can be justified requires further consideration of the PI time savings. In accordance with numbers of missions and numbers of targets covered per mission in the Input/Output

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

TABLE XI

PI PRESENTATION CHARACTERISTICS

WITH THE HEADER AND SETS 1 THROUGH 4 GIVEN ROUTINELY:	FIRST PHASE	SECOND PHASE
Probability of Requiring no more sets	.812	.855
Probability of Requiring one or more sets	.188	.145
Probability of Requiring two or more sets	.013	.008
Expected Number of Requests/Target	.200	.152
Expected Number of Characters Requested/Target	1510	488
Normalized Time Devoted to Data		
if r = 0.1	.192	.383
if r = 0.03	.178	.373
if r = 0	.172	.368

from Table XI and 20 minutes for first phase and 6 minutes for second phase as the time required if all the data were given, the manpower saving is 1630 manhours per month. Thus if PI time is valued at more than [redacted] per hour, the cost of the printers can be justified. 25X1A

The provision of the request capability with the printers as described above will have two additional benefits. First, the printers could be used to prepare the Presentation Sheet right at the PI location and thus eliminate the manpower currently required to distribute this material. Second, since less than a third of the total data available is printed out, there will be less classified material to gather up and destroy.

5.3 First and Second Phase Studies

In an effort to ascertain the sensitivity of NPIC to variations in manpower loading, a simple model of First and Second Phase Readout activities was prepared. This model was then utilized as the basis for studies and analyses of costs, time, personnel and reliability as related to the immediate readout activities. The two sections that follow discuss aspects of the above parameters and the knowledge gained as a result of this effort.

5.3.1 Cost vs Time

This section discusses the results of certain systems studies which were made to ascertain the variation of costs and time that would result if different numbers of people were utilized during First and Second Phase Reporting. The objective of the study was to identify any unexpected sensitivities of cost or time to variations of the parameters considered and to help in the further selection of detailed design parameters. The operational model of the partial system discussed in Appendix E was utilized for these studies and proved to be of sufficient detail to satisfy the desired objectives.

The study examined, from a personnel loading point of view, first phase and second phase reporting independently and then combined the independent results and added cost information as a final effort. These points of view are examined in more detail in the sections that follow.

5.3.1.1 Range of Parameters

Table XII has been prepared to illustrate the ranges over which the various parameters were examined in the phase of the study effort. The results of certain parts of the study are somewhat sensitive to the assumptions made for labor costs and therefore these costs are defined below.

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

TABLE XII

RANGES OF PARAMETERS

VARIABLE	HIGHEST PRIORITY TARGETS	OTHER TARGETS	COST
PI - Time/Target	90 - 514 minutes	30 - 129 minutes	
Data Entry/Target	4 minutes	.8 - 4 minutes	
Verify/Target	10 minutes	.2 - 10 minutes	
Edit/Target	4 minutes	.8 - 4 minutes	
Approve/Target	2 minutes	.4 - 2 minutes	
Average Number of Targets/Bucket	400 -1000	2600	-
All Consoles			$100/month

* Labor Costs are estimated as costs per productive hour and consider Basic annual salary, fringe benefits and productive hours/year.

25X1A

TOP SECRET

TOP SECRET

25X1

25X1

SCR 352

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

Cost per Production Hour

25X1A These costs are derived by modifying the basic assumed annual salary of the various skills to account for productivity and fringe benefits. For example, the basic assumed annual salary of PI was assumed to be [redacted] To this was added a 25% fringe benefit, raising the annual cost 25X1A to [redacted] Dividing by a productive capability of 1352 hours per year results in a productive hour cost of [redacted] The other labor costs were 25X1A estimated by the same technique.

Assignment of Costs

In assigning labor costs to the analysis, the assumption has been made that all persons needed in the First Phase Readout operation are assigned for the entire operation and similarly all people needed for Second Phase are assigned for the entire time required to accomplish the Second Phase Readout.

5.3.1.2 First Phase Only

Figure 21 shows the average number of personnel required to read out and process 400 targets in differing total amounts of time ranging from hundreds of minutes to approximately 6 working days. The solid lines represent the numbers of PI's needed; the dashed lines represent the remaining personnel (referred to as "support personnel" for the sake of simplicity). The curves are annotated to show the average PI Time/Target.

The number of support personnel in each category can be derived by assuming that for every ten, the division of responsibility will be as follows:

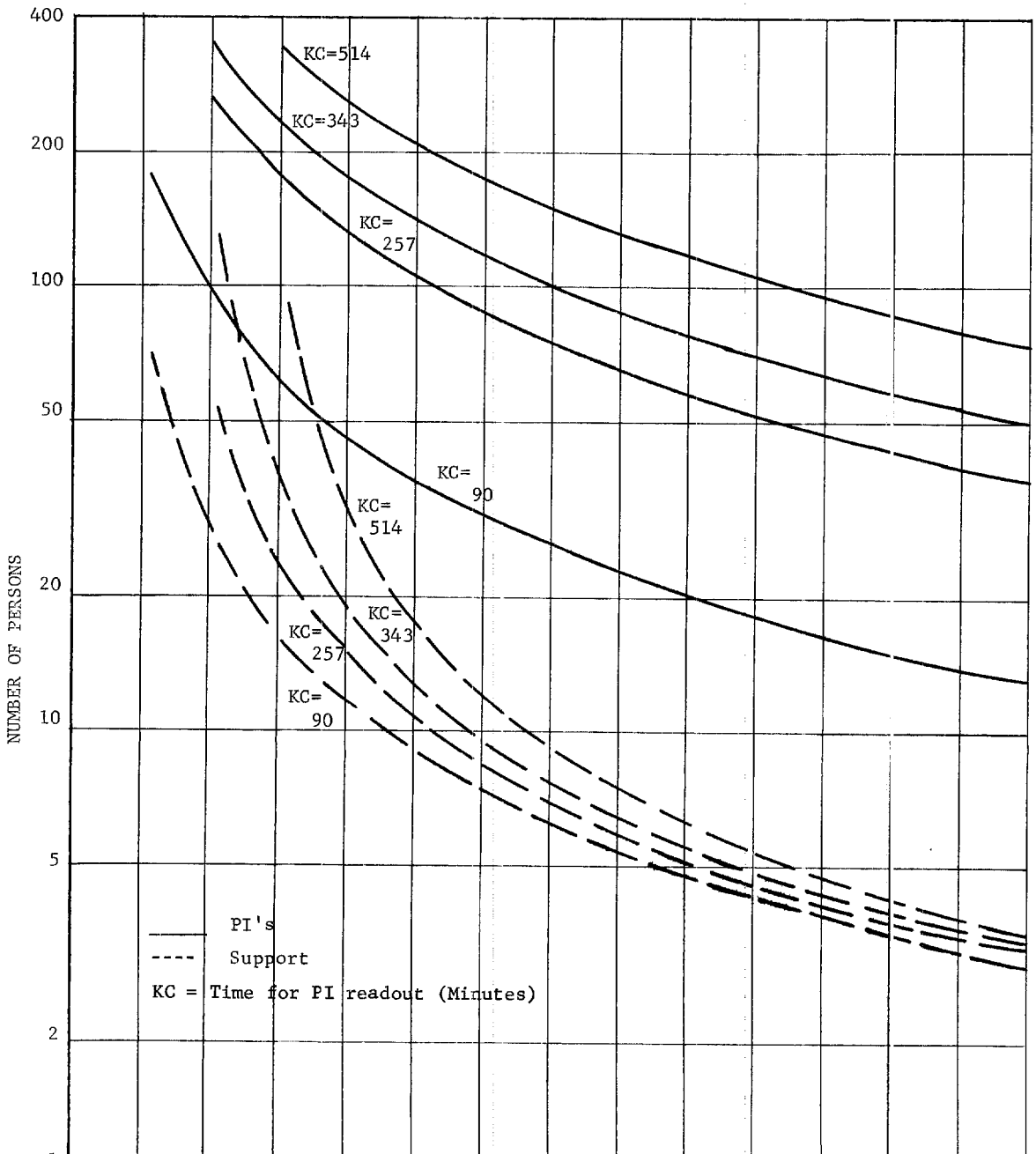
Data entry	2
Verify	5
Edit	2
Approve	1

25X1

TOP SECRET

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1



For example, if the curve shows the need for 20 support people the ratios when applied would yield:

Data Entry	4
Verify	10
Edit	4
Approve	2
	20

The curves in general follow the anticipated inverse relationship between time and people, that is, the more time allowed the fewer people required. The interesting point brought to light by this effort is that abnormally large numbers of support people are needed to process the readout material if the allowable time for readout is short. In general it appears to be unwise to schedule completion of first phase readout in time periods less than 2.5 times the average time that the PI utilizes in the readout of a target. (Average taken over the whole set of readout targets).

5.3.1.3 Second Phase Only

Figure 22 shows the number of people needed to read out and process 2600 targets in differing total amounts of time ranging from hundreds of minutes to approximately 6 working days. Once again the solid curves represent PI's and the dashed curves represent support personnel. The ratio of skills in the support area is the same as in the previous section. The time required by the support prople to process the readout data has been decreased by a factor of 5 because of the anticipated small effort required to process each individual readout.

Two interesting points are observed in these curves. The first is that the number of support people required is only slightly influenced by the average time required by the PI to perform the readout of a target and secondly that no abnormal numbers of support people are required if the time to readout and process 2600 targets is compressed.

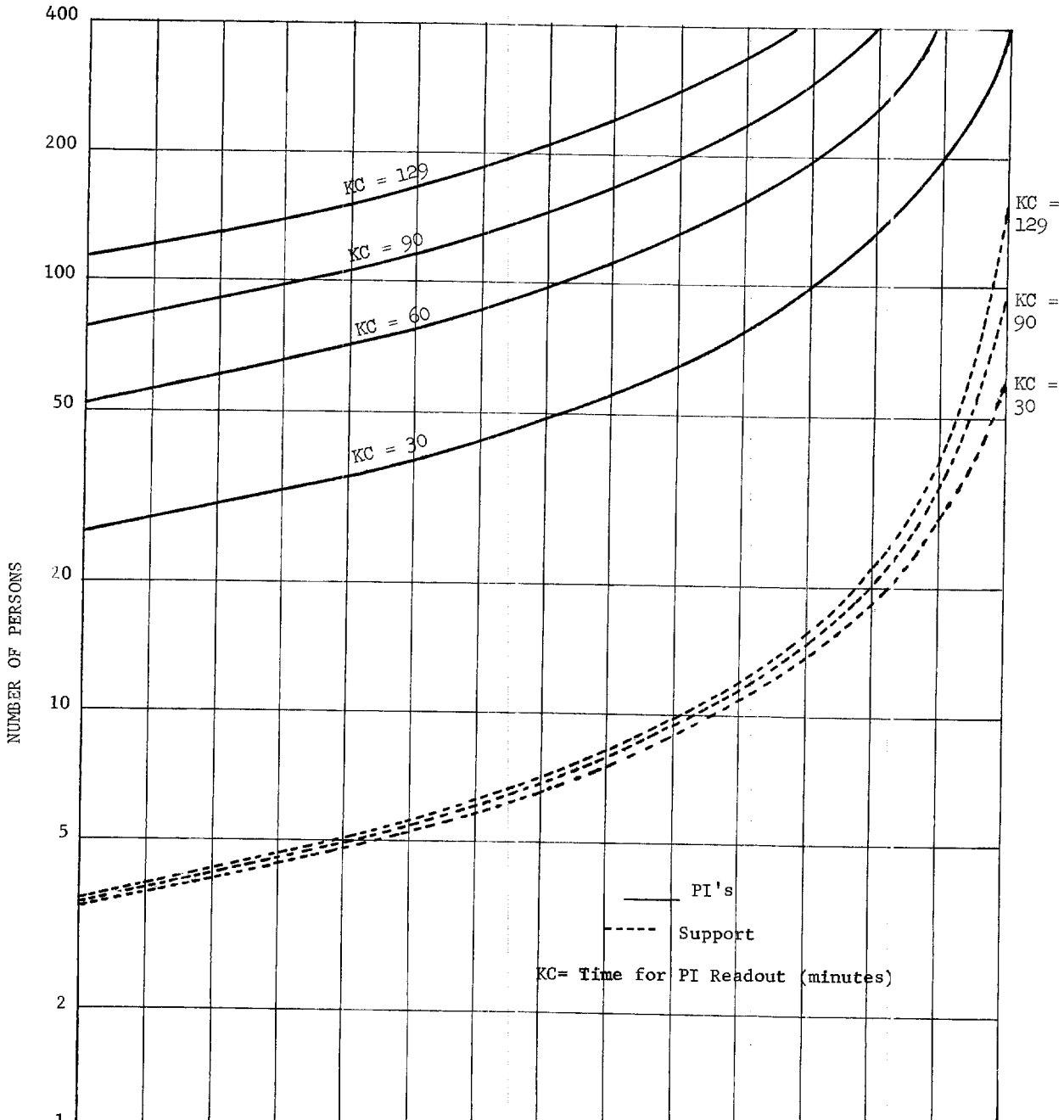
It should be noted that the time scale on Figure 22 has been reversed to permit overlaying with Figure 21.

25X1

TOP SECRET
Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

SCR 352



required for both First and Second Phase Readout and will be invariant for a fixed horizontal transition of the figures. This technique was utilized to permit further analysis of the basic data. In particular it was relatively simple to obtain system manning data pertaining to a condition where upon the same numbers of PI's were to be utilized in both First and Second Phase Readout. The results of this effort are discussed in the next section. The section following that considers the case where the same numbers of support personnel have been assigned to both First and Second Phase Reporting. An estimated 12000 First and Second Phase Readouts per month have been used in generating the cost figures.

Figure 23 has been prepared to illustrate the monthly labor and console costs that can be expected if the number of PI's is the same during First and Second Phase readout. The numbers of PI's required to meet the various scheduled readout times are shown on the abscissa. The average time required by a PI to readout First and Second Phase targets is taken as 90 minutes. (The bands indicate the variations as the First Phase readout time varies between 90 and 480 minutes.) The graph on the left includes
25X1A the cost of providing consoles to both PI's and support personnel while that on the right provides consoles to support personnel only. This data illustrates that costs do not vary greatly as long as the scheduled time is greater than four days, the variation being from about [redacted] to [redacted] per month on the left and [redacted] on the right. 25X1A 25X1A

Figure 24 shows the same type of costs when the number of support personnel, rather than the number of PI's is the same during First and Second Phase readout. The values of PI readout time are also the same. Again the variation is small and in the same range as above as long as the scheduled readout time is greater than four days.

Other cases have been investigated and they all tend to confirm that costs (other than PI console costs) for First and Second Phase Readout are only a function of the average time that a PI spends doing readout of Targets. Figure 25 illustrates this fact for average readout times of 1.5, 2 and 3 hours.

5.3.1.5 Conclusions and Recommendations Pertinent to System Design

The effort reported on in this section leads to the recommendations that First and Second Phase readout be considered for design purposes as being accomplished over a time span of 5 days. It appears that the

25X1A

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

Next 2 Page(s) In Document Exempt

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

the system rather insensitive to values selected in estimating the various parameters.

5.3.2 Console Reliability

The model described in Appendix E was used in a study of the effects of console reliability upon the delivery of First or Second Phase Products. In the specific cases investigated the numbers of targets to be included in the product was taken to be 400 and 1000, the average time utilized by the PI was taken to be 90 minutes and two hours. The assumption was made that the mean time between failures of a console was 1000 hours. Furthermore the failure of a console during the preparation of a product would negate the use of that console for the entire time of preparation of the product. For the purposes of this study, it was further assumed that there would be 30 products a year prepared. The results of this study are shown on Figure 26 which also shows the lines of constant lateness.

The narrow spread of results with the discussed variation of parameters lead to a more detailed investigation to ascertain the reason for the apparent insensitivity. The recording of this effort is shown in Appendix F and the application of the results are also recorded on Figure 26.

In the One Process Case, we have considered the four operations (data entry, verify, edit and approve) in the support area as one large process which allows complete interchangeability of consoles. In the four process case, we have considered the four support operations as separate processes. As determined in Appendix F the mathematics of the problem shows that the effect of console failures on product delivery is dependent only upon the time required to prepare the product and the Mean Time between Failures (MTBF) of the equipment. It is evident from the figure that if the design time to deliver the product is decreased, percentage wise less time is lost.

The obvious conclusion derived from this study is that there are significant advantages to be gained, in operations involving processes of the type considered, by making consoles similar and in allowing interchangeability. Furthermore, if repairs of the consoles must be postponed until the completion of the product generation, it is advantageous to commit additional resources to the preparation of the product. In view of the short amount of time lost because of console failures, it appears reasonable to schedule products a few minutes later rather than to supply backup consoles.

5.4 System Capacity and Queues

An analysis has been made to examine the extent of queues under various operating loads. The basic analysis was performed utilizing the

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

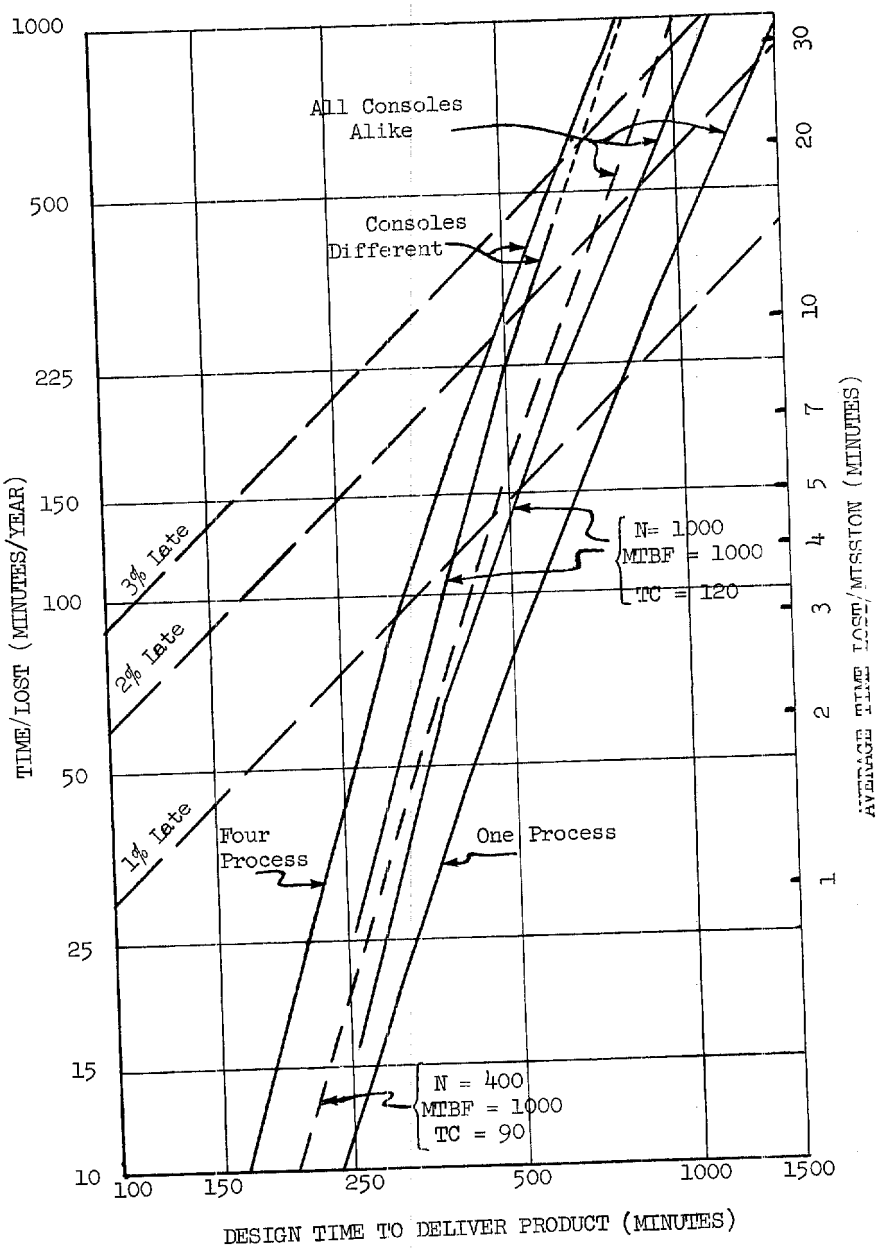


Figure 26 Product Time Lost/Year
as a function of time required to produce product

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

request rates developed in Appendix B and the material presented as Appendix G and examining various operating modes.

There are basically two different modes in which the operations of the central processor can be viewed. In the first, each of the 309,000 data base accesses/month (See Appendix B) is considered as an independent event and it is assumed that 250 milliseconds is required to service each access. (This is equivalent to 2.5 accesses to a large drum and makes provision for properly identifying, locating, and retrieving a piece of data. It allows 10 milliseconds for data transfer and processing that does not overlap drum access.) This mode of operation corresponds to the lower curve on Figure 27. Assuming a single shift, five-day week rate of 309,000 accesses/month and the service time of 250 milliseconds establishes a load factor of 0.125 and thus the point on the curve designated as the "monthly average." The worst day is based on data base accesses for 500 First Phase readouts, 500 Second Phase readouts, 60 Third Phase readouts, and 2000 COINS queries. This produces a load factor of approximately 0.23 and establishes the point designated at "worst day." The worst hour is based on doubling this load factor. Thus when viewed in this mode, the mean number in the queue is between 0.01 and 0.20, and the mean time in queue is between 0.02 and 0.10 seconds.

The other mode in which operation may be viewed is developed by assuming that the 309,000 accesses are a result of a lesser number of queries where each query requires several pieces of data. The case that was investigated assumed an average of 5 accesses per query and a Poisson distribution of the number of accesses per query. This mode of operation is described by the upper curve on Figure 27. The same load factors for monthly average, worst day, and worst hour apply, resulting in a mean time in queue between 0.19 and 1.05 seconds. The actual operation of the central processor will probably fall somewhere in between the two curves; the mode described by the upper curve is used in the remainder of this analysis since it represents a conservative position.

Even though the cases above do not represent long queues, the advantages of assigning different priorities were investigated. The data presented in Table XIII and Figure 28 illustrates the time in queue of queries under various conditions of loading and in the case where certain categories of operations are assigned a lower priority for service. In the example utilized in preparing this table it was assumed that the average query would require 1.250 seconds to be serviced (i.e., 5 accesses to data at 250 milliseconds each). As can be seen, even under the worst case, the average time in queue for the lowest priority query is only slightly larger than the average service time.

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

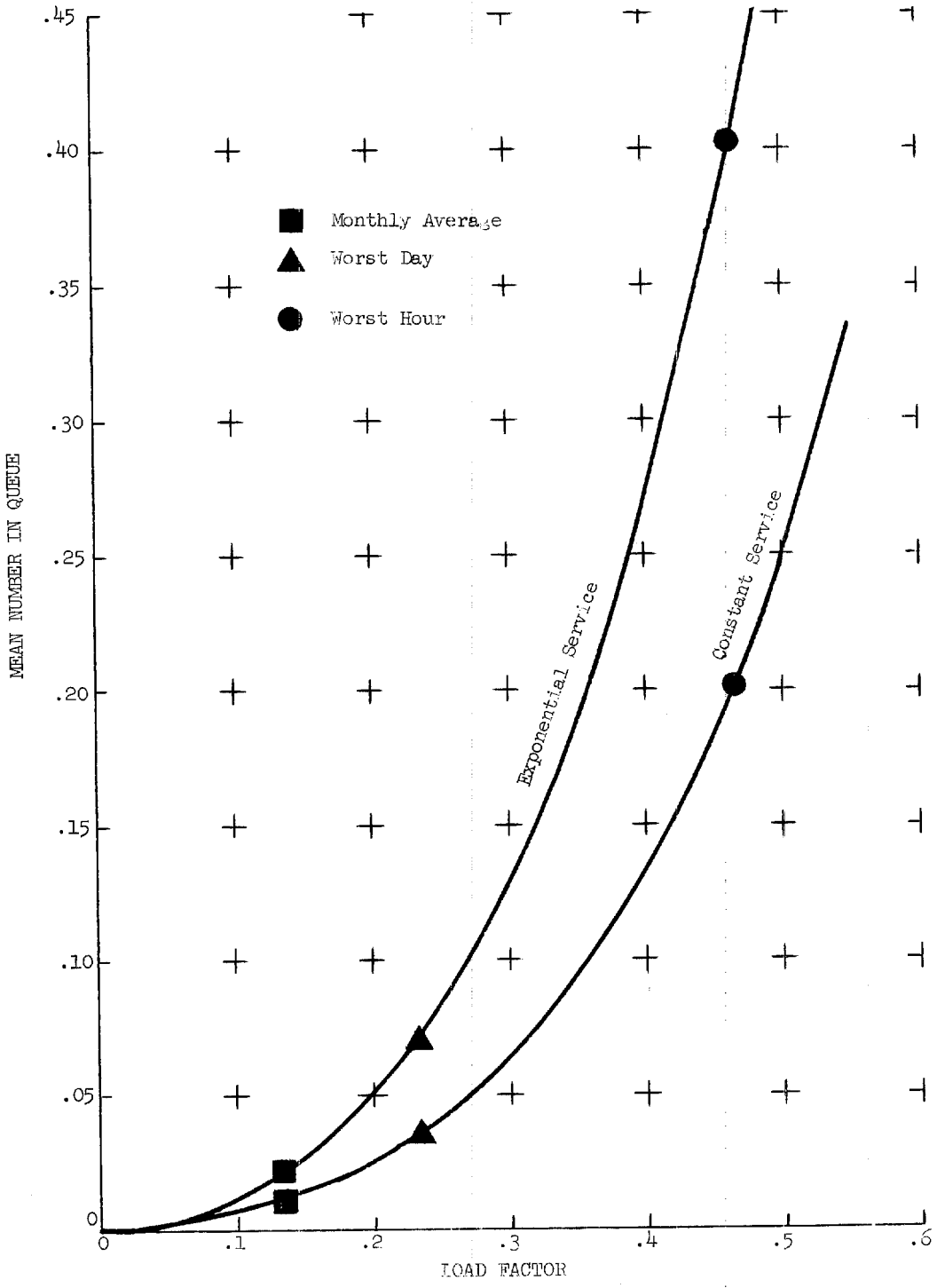


Figure 27 Expected Queues for Storage Access

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

TABLE XIII

AVERAGE WAITING TIME IN QUEUE

	No Priorities	UPDATING SECOND PRIORITY		COINS - SECOND PRIORITY	
		Priority I	Priority II	Priority I	Priority II
Worst Hour	1.05 sec.	0.906 sec.	1.645 sec.	0.777 sec.	1.411 sec.
Worst Day	0.352 sec.	0.348 sec.	0.449 sec.	0.326 sec.	0.421 sec.
Average Monthly	0.193 sec.	0.188 sec.	0.217 sec.	0.1835 sec.	0.212 sec.

TOP SECRET

TOP SECRET

25X1

25X1

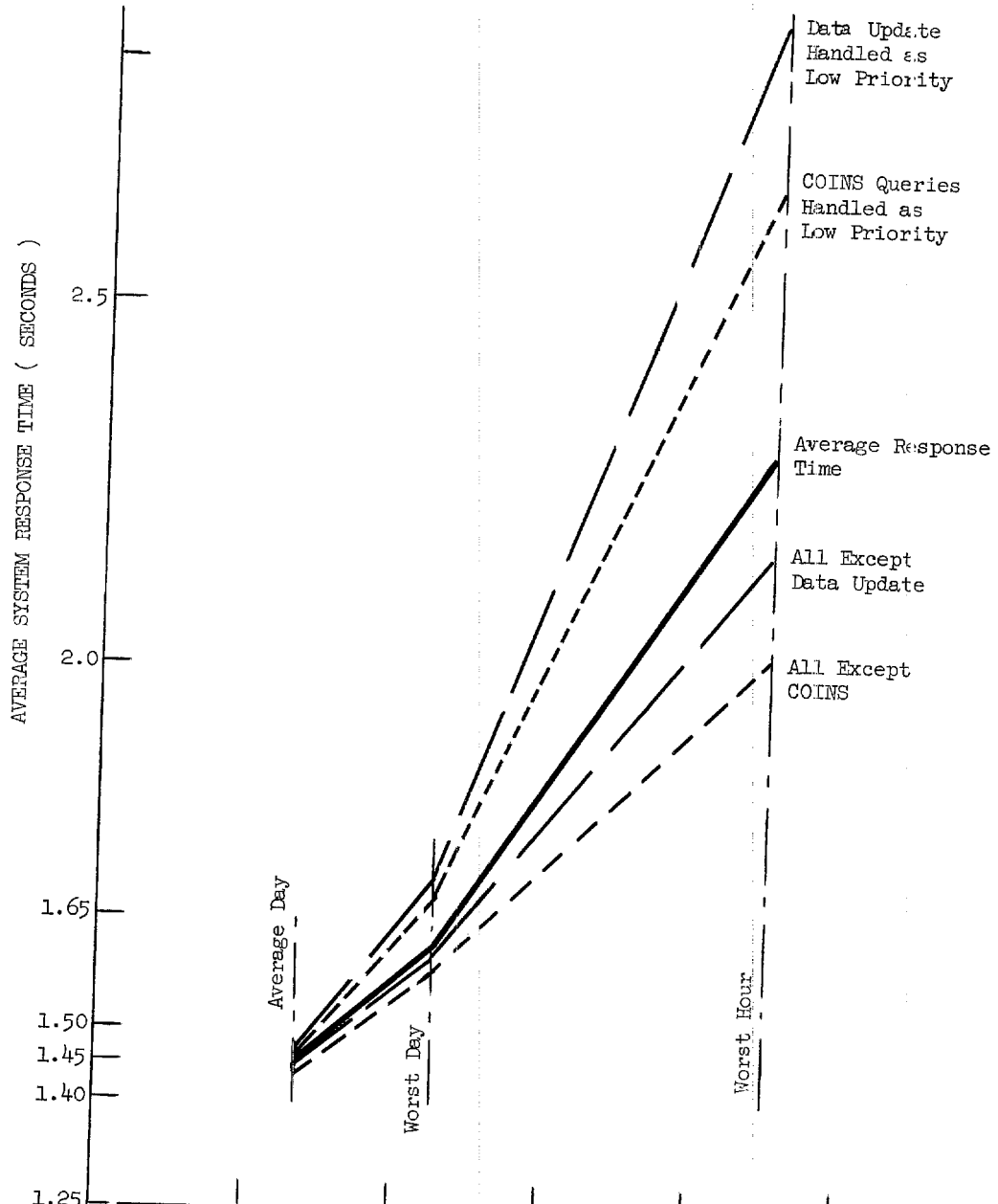
Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

TOP SECRET
Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

SCR 352



Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

In conclusion, it appears that the loads and response times of the system will not cause any queuing problems even if all operations are performed during the first shift, including data base updating functions, and gathering of data for presentations.

5.5 Input/Output Console Study

A study was undertaken to ascertain the numbers and types of input/output consoles that should be in the recommended system configurations. The study was broken into basically two different parts. The first part was concerned with the determination of the types and numbers of consoles that were necessary in order to satisfy the basic system design. The second part of the study was to ascertain if the numbers selected were sufficient in view of the mission aspect of system loading. The results of the two parts of the study are contained in the following section.

5.5.1 Determination of Console Types and Numbers

An analysis was undertaken to determine the total utilization of Input/Output consoles in the IIS. The activity charts (Section 3 of Volume 2) were examined to determine the processes which have been designed to include man/machine interfaces. These interfaces were identified and the required characteristics, utilization factor and Center location of all of the identified consoles was recorded.

The required characteristics were analyzed to ascertain any patterns and it became apparent that two general sets of equipment would probably be sufficient in satisfying the identified characteristics. Table XIV was prepared to illustrate these equipment groupings.

The underlying characteristic of the Set I equipment was the fact that they would be utilized primarily by specialists who would be familiar with and indeed operate the equipment. The equipment therefore should exhibit rapid input/output characteristics and should permit time saving operations where possible.

The Set II equipment on the other hand would be utilized in places and operations where personnel would interact less intimately with the system and where somewhat slower response time would not seriously hamper the operation being supported. It was recognized however that in general printers should be capable of operating at speeds equal to or in excess of reading speeds.

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

TABLE XIV

IIS CONSOLE CHARACTERISTICS

Console Type	Quantity	Keyboard	CRT Display	Cursor	Special Function Keys	Upper/Lower Case A/N
Group I						
PD Edit	6	x	x	x	x	x
PAG Edit	3	x	x	x	x	x
CSD Edit	3	x	x	x	x	x
PAG Approval	2	x	x	x	x	x
PAG Verification	7	x	x	x	x	x
CSD Verification	2	x	x	x	x	x
PAG Query (SUSIE)	5	x	x			
CSD Query (SUSIE)	2	x	x			
PAG Data Entry	4	x				x
PD Data Entry	3	x				x

	Quantity	Printer	Keyboard	Tape Punch & Reader		
Group II						
OS TTY	1	x	x	x		
PAG PI-1	10	x	x			
PAG PI-2	10	x				
CSD	2	x				
IAS	1	x				
DIAAP	1	x				

TOP SECRET

TOP SECRET

25X1

25X1

SCR 352

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

reliable and also more flexible. (See Section 5.3.2). The PAG Query Console and the CSD Query Consoles were identified as a similar pair as were the PAG and PD Data Entry devices.

It is obvious that the Set II equipment also could be broken down into three distinct equipment types.

The utilization figures were then analyzed together with location data and the numbers of consoles to be recommended for each location selected. This selection was previously identified in Table VI of this volume.

5.5.2 Analysis of Console Utilization

A hypothetical mission readout scheduling chart was prepared which illustrated the readout of 3 satellite missions and numerous aircraft missions. This scheduling chart considered a 4 day, a 5 day and a 6 day satellite readout together with sufficient 100 target aircraft missions to make up the anticipated 13,200 target readouts per month. In keeping with JIIRG assumptions and the 20% reduction in PI readout time, High Priority Targets were scheduled for 4.8 hours average readout time and low Priority for 1.2 hours average readout time. Personnel were assigned to actual readout and to support functions in the same manner as previously discussed (see Section 5.3.1.4). The number of PI's and the numbers of support personnel necessary are shown on Table XV and XVI.

The total numbers of PI's (224) was selected from the projections made and discussed in Section 6. The numbers of consoles selected was of course made by summing up the Data Entry, Verification, Edit and Approval consoles assigned to the PAG area.

The number of PI's involved in Phase III (Detailed Reporting) is the difference between those required for the Phase I and Phase II effort and the total of 224. The number of support consoles available for Phase III was similarly obtained from the total of 16 consoles.

Table XV considers the case where the assignment of personnel to First and Second Phase readout of satellite missions has been done in such a way as to make the number of support people the same for both phases. (See for example Section 5.3.1.4). It is evident from the table that 16 consoles are sufficient to handle all First Phase and Second Phase

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

TABLE XV

HYPOTHETICAL MANPOWER SCHEDULING (CONSTANT SUPPORT)

DAYS	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
PHOTO INTERPRETERS																					
Phase I KH Missions	112	112					78	78	78						94	94	94				
Phase I - Other			20		20	20		20		20	20		20	20		20	20		20		20
Phase II - KH			196	196						122	122	122					76	152	152		
Phase II - Other	20	20		20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20
Phase III	92	92	8	8	184	184	126	106	126	62	62	82	184	184	110	90	61	52	32	204	184
TOTAL	← 224 →																				
* SUPPORT PERSONNEL (WITH CONSOLES)																					
Phase I - KH	12.7	12.7					7.6	7.6	7.6						9.4	9.4	4.7				
Phase I - Other			2.3		2.3	2.3		2.3		2.3	2.3		2.3	2.3		2.3	2.3		2.3		2.3
Phase II - KH			12.7	12.7						7.6	7.6	7.6					4.7	9.4	9.4		
Phase II - Other	1.3	1.3		1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3
Phase III **	2.0	2.0	1.0	2.0	12.4	12.4	7.1	4.8	7.1	4.8	4.8	7.1	12.4	12.4	5.3	3.0	3.0	5.3	3.0	14.7	12.4
TOTAL	← 16 →																				

* Support Personnel are limited to Data Entry, Verification, Edit and Approval only.

** Number of Support Personnel at Consoles available in PAG Area To Support Phase III Activities

TOP SECRET

TOP SECRET

25X1

25X1

SCR 352

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

TABLE XVI

HYPOTHETICAL MANPOWER SCHEDULING (CONSTANT PI's)

DAYS	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
PHOTO INTERPRETERS																					
Phase I	153	76					102	102	51						122	122					
Phase I - Other			20		20	20		20		20	20		20	20		20	20		20		20
Phase II		77	153	153					51	102	102	102					122	122	122		
Phase II - Other	20	20		20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20
Phase III	51	51	51	51	184	184	102	82	102	82	82	102	184	184	82	62	62	82	62	204	184
TOTAL	← 224 →																				
SUPPORT PERSONNEL (WITH CONSOLES)																					
Phase I	21.8	10.9					10.7	10.7	5.4						14.7	14.7					
Phase I - Other			2.3		2.3	2.3		2.3		2.3	2.3		2.3	2.3		2.3	2.3		2.3		2.3
Phase II		4.8	9.6	9.6					3.1	6.2	6.2	6.2					7.5	7.5	7.5		
Phase II - Other	1.3	1.3		1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3
Phase III	-7.1	-1.0	4.1	5.1	12.4	12.4	4.0	1.7	6.2	6.2	6.2	8.5	12.4	12.4	0	-2.3	4.9	7.2	4.9	14.7	12.4
	← 16 →																				

TOP SECRET

TOP SECRET

25X1

25X1

SCR 352

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

Table XVI investigates the same hypothetical month where the assignment strategy for satellite missions has been to assign the same numbers of PI's to both First and Second Phase Reporting. As can be seen there is a deficiency of consoles available on days 1, 2 and 16 when a total of only 16 consoles are available. These deficiencies occur on days of First Phase Readout of satellite missions and are brought about primarily by the manner in which the model utilized in Section 5.3.1.4 was designed. In real practice a similar situation would occur and would be brought about by the fact that on the first day of an OAK report, large numbers of complex readouts would not be delivered to the support team until late in the day. Consequently, abnormally large (compared with the rest of the readout) numbers of support people would be needed for short periods of time.

5.5.3 Conclusions

It is concluded as a result of this study and analysis that 6 different console types should suffice to support the I/O man/machine console requirements of the IIS. They are Edit consoles with upper and lower case capabilities, (including verification, edit and approval), the DEN console (Data Entry), the Query Console (previously identified as SUSIE console in SCR 288), the Print Punch and Keyboard device, the Print and Keyboard device and the Print Only device.

It is further concluded that 16 consoles are probably sufficient for use in satisfying the Data Entry, Verification, Edit and Approval function of First and Second Phase reporting in all cases except those in which more than 120 PI's are assigned to do First Phase Readout of a Satellite Mission.

Changes in scheduling policy from constant PI's to Constant Support could alleviate this problem as well as additions to the numbers of consoles. Until more accurate estimations can be made of the actual time it takes to read out High Priority COMIREX Targets, it seems reasonable to consider the selection of 16 consoles in this activity area as sufficient.

The operations in other areas using consoles are not as critical as those in the PAG area and therefore a more thorough analysis of these operations did not appear necessary.

5.6 Language Considerations and Alternatives

5.6.1 Query Language Considerations

use. This query capability is brought about by the implementation of a query language. The specific capabilities of any query language are a function of the system user, i.e. the capabilities should be tailored to meet the requirements of the user. This means, that although there may be a degree of commonality among various query languages, one must be concerned mainly with the satisfaction of user needs in selecting the one appropriate to a particular situation. The specific capabilities of the query language required for NPIC are given in Section 4.5.2.3. In considering the alternatives for meeting these requirements, a major consideration must be the retrieval of information for timely use by dialogue. To support a dialogue, the timing should be on the order of seconds, the normal attention span of a searcher, not several minutes or hours.

5.6.1.2 The Alternatives

As was stated in paragraph 5.6.1.1 existing query languages do, in fact, have a degree of commonality, i.e. some existing query languages could possibly meet the capabilities and timing considerations discussed above. This fact leads to the alternatives of (1) implementing a new query language for the IIS, (2) utilizing the Technical Information Processing System (TIPS) developed at NSA, or (3) implementing a query language similar to the one that is part of the CAPRI system being developed at CIA. Each of these three alternatives has been examined from the standpoint of

- Capabilities
- Known advantages
- Known disadvantages
- Estimated cost to implement at NPIC under the IIS

Following the discussion of the three alternatives, we have provided our conclusions and recommendations.

5.6.1.2.1 CAPRI at CIA

CAPRI is defined as the ADP software being developed to support Project CHIVE at CIA. Project CHIVE is a long range system concept which encompasses the design and installation of a total system including procedures, organization and equipment. CAPRI then, includes software for

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

designing CAPRI have not been able to obtain detailed specifications for CHIVE. This generalized approach (highly structured data base) is also necessary when the specific use of a data base is not known. In turn, because of the highly structured data base, both the file maintenance program and the query language become exceedingly sophisticated and generalized.

Therefore, on a design level (though not on the coding level) both the query language and the file maintenance routines could be utilized at NPIC. Obviously, implementation of the total design of these two software packages would be unnecessary, because the IIS data base as conceived is less highly structured (more free text sectors) than that of the CAPRI data base. Therefore, some subset of the program design for CAPRI could be useful in the program design of the query language and file maintenance programs for the IIS.

a. Advantages

1. The scope of CAPRI is broad enough that it probably covers (in a design sense) the NPIC query language and file maintenance functions.

2. A common data base format, query language format and maintenance program design between Project CAPRI and the IIS would result in a software design savings for both systems for any future software design efforts, e.g. extended capabilities to the query language.

b. Disadvantages

1. As CAPRI is being implemented on IBM 360 equipment, all coding effort will be of little or no use to the IIS (reprogramming effort would represent nearly the same cost as implementing a new system for the IIS).

2. The CAPRI system is highly generalized which in turn necessitates incorporation of inefficiency and extra cost into the IIS. This inefficiency and extra cost are created by the fact CAPRI is generalized to

3. The query language capabilities of CAPRI are not formally documented at this time which makes it difficult to extract and evaluate possible design features that could be utilized in the design specifications of the IIS. (This is the case because CAPRI is in the implementation stage and sufficient design documentation will not be available until after implementation).

c. Estimated Cost to Implement at NPIC

The estimated cost to implement a subset of the CAPRI system would be higher than implementing a new query and maintenance capability for the IIS. This is the case because of the generality of design of the CAPRI software, and the extra formatting of the data base. How much more costly a CAPRI type implementation would be is difficult to estimate but as a minimum a 20% extra cost figure is probably conservative. Therefore, as a conservative minimum, the estimated cost to implement a CAPRI type system (query and maintenance only) for the IIS would be in excess of 12 man-years.

5.6.1.2.2 TIPS at NSA

TIPS is a programming system developed at NSA which gives the user both a query and data maintenance capability. The present operating version of TIPS is known as "TIPS PILOT". This version of TIPS is designed to accommodate a serial access (tape-stored) data base, where the search analogue for a query must be completely expressed before a tape search is made on a match/no match basis. TIPS PILOT consists of five programs known as TIPS elements:

- STAGING - develops query analogues
- ROUTER - provides a control function
- SEARCH - generalized tape search program that utilizes the query analogue to extract "match" records
- CLEANUP - brings together all extracted data resulting from the search element
- REMAKE - controls file maintenance functions

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

are:

- Query - (JAKEA) programs - query programs written for each particular file and for the particular type of query function

- Result Worker- (JAKEB) programs- programs to reformat and summarize the extracted information.

Therefore, any consideration of implementing a TIPS PILOT system for the IIS must take into account the fact that a large portion of the programming (query and result worker programs) would have to be created for the IIS.

a. Advantages

1. A good portion of a query and maintenance system are available at no initial cost from NSA because of the hardware (Univac 490-494 equipment) similarities between NSA and NPIC.

2. A limited query capability on the Target Brief File (TBF) could be implemented before the IIS is operational. (This indeed will be implemented, in part, for the COINS I experiment at NPIC).

3. Any additional development of TIPS at NSA could be utilized at NPIC at no initial cost. This additional development would take the form of TIPS-I or TIPS Interrogation Language (TILE)

b. Disadvantages

1. No individual(s) at NPIC has an intimate knowledge of TIPS system and therefore maintenance at NPIC of the TIPS system would tend to run higher in cost (estimated conservatively at 10% higher) than a system developed from its inception by the Center.

2. Response time to queries is

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

c. Estimated Cost to Implement at NPIC

The estimated cost to implement a TIPS PILOT system (including necessary query and result worker programs) is 5.2 man years.

5.6.1.2.3 PIRL at NPIC

PIC Information Retrieval Language including the file maintenance capability is a portion of the software design for the IIS and is described in detail in Volume 2 - System Design Description - Section 5.2.

a. Advantages

1. Obviously PIRL is being designed to meet all the capabilities and response time stated in Section 5.6.1.1.

2. This approach eliminates the disadvantages discussed under CAPRI and TIPS systems.

3. An integrated development effort of all IIS software can eliminate redundant programming, i.e., two search routines that perform the same function.

4. The documentation standards of all IIS software would comply to NPIC standards.

5. The growth (new design features to the software) would not depend on an outside agency who is not directly concerned with NPIC problems.

b. Disadvantages

1. Initial implementation cost will be higher than that of a TIPS system.

2. A TIPS system could possibly be implemented sooner than the IIS system.

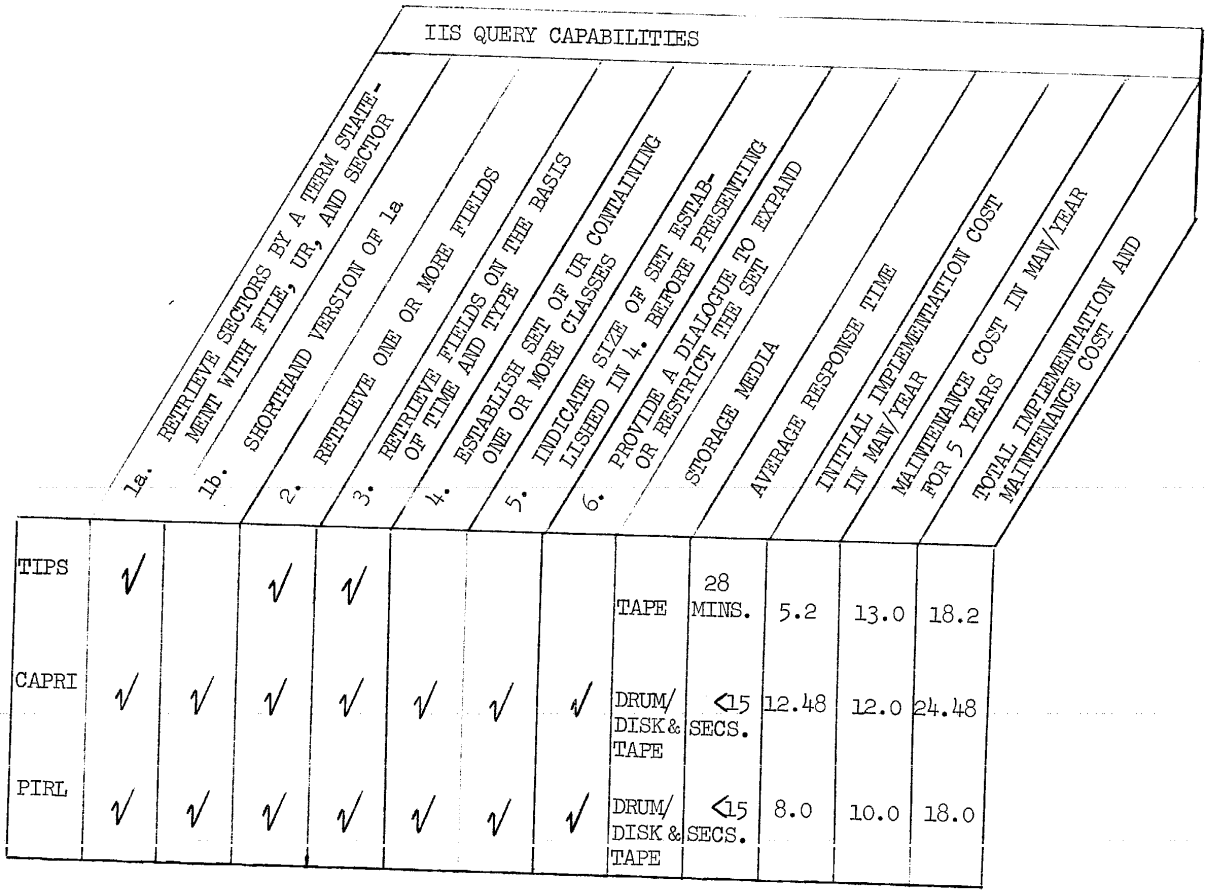
c. Estimated Cost to Implement at NPIC

The estimated cost to implement PIRL is 8.0 man years.

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

TOP SECRET



	IIS QUERY CAPABILITIES										
	1a. RETRIEVE SECTORS BY A TERM STATEMENT WITH FILE, UR, AND SECTOR	1b. SHORTHAND VERSION OF 1a	2. RETRIEVE ONE OR MORE FIELDS	3. RETRIEVE FIELDS ON THE BASIS OF TIME AND TYPE	4. ESTABLISH SET OF UR CONTAINING ONE OR MORE CLASSES	5. INDICATE SIZE OF SET ESTABLISHED IN 4. BEFORE PRESENTING	6. PROVIDE A DIALOGUE TO EXPAND OR RESTRICT THE SET	STORAGE MEDIA	AVERAGE RESPONSE TIME	INITIAL IMPLEMENTATION COST IN MAN/YEAR	MAINTENANCE COST IN MAN/YEAR FOR 5 YEARS	TOTAL IMPLEMENTATION AND MAINTENANCE COST
TIPS	√		√	√				TAPE	28 MINS.	5.2	13.0	18.2
CAPRI	√	√	√	√	√	√	√	DRUM/ DISK & TAPE	<15 SECS.	12.48	12.0	24.48
PIRL	√	√	√	√	√	√	√	DRUM/ DISK & TAPE	<15 SECS.	8.0	10.0	18.0

Figure 29 IIS Query Capabilities

25X1

TOP SECRET

25X1

SCR 352

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

a. The capabilities as expressed in 5.6.1.1 as to whether each approach meets or does not meet the capability.

b. Storage media the system is designed for, e.g. serial access or random access.

c. Average response time of each system to a query (this is defined as the time lapse between completed keyed in query and beginning of response to the requesting user.)

d. Estimated cost to initially implement each system into the IIS.

e. Estimated cost to maintain (includes corrections and adding capabilities to the software) each system over a period of five years.

Figure 29 then shows the conclusion that the PIRL system gives NPIC the most capability for the least cost.

5.6.2 Program Language Alternatives in IIS Implementation

Past programming efforts at NPIC have all been coded in one of two assembly languages i.e. 490 - SPURT and 494 - SPRYE. The SPRYE 494 Assembly Language developed at NSA is presently operational at the Center. Therefore there are four major alternatives in selecting the coding tool for IIS implementation: (1) all IIS coding to be done in SPRYE, (2) all IIS coding to be done in the general-purpose compiler language, e.g. COBOL, FORTRAN, (3) all IIS coding to be done in a new specialized compiler language, or (4) a mix of coding, some programs in assembly language, others in a compiler language.

A new, specialized compiler language, called the Systems Language is described below. This is followed by a comparison of the various alternatives and the resulting conclusions.

5.6.2.1 Systems Language

The development of the System Language began at [redacted] some months before the current IIS effort was started. The genesis of the development effort was the continuing failure of the computer industry to develop a simple, general purpose language oriented to information processing and

25X1A

The first factor is the so-called "information explosion". Primarily because of the advances in computer hardware technology and the resultant reduction in hardware costs, automated techniques are being applied to the collection, utilization, and dissemination of information. Information may have a time value such as "latest available" as is the case of NPIC. Through mass storage devices, the information or pertinent portions of information can be presented to users in a matter of seconds. The convenient manipulation of information can no longer be ignored, and systems must be designed to perform this manipulation in a timely manner.

The second factor is the steadily rising cost of the personnel required to create an information system. As the personnel costs rise, it becomes more urgent to provide tools for more productive work from each individual. As a new system begins to develop, a vast array of unpredicted functions are proposed for inclusion, placing strains on most manpower projections. The demand can be met by either more personnel or more productive use of existing personnel. Obviously, under budgetary constraints, the latter alternative is more reasonable.

Many computer languages have been proposed, developed, and implemented over the years but only one (FORTRAN) has a real claim to universal usage. This language has served a very real need to the scientific community. Today it remains as a very simple mathematical language even with the refinements of three dimensional arrays, matrix manipulation, etc. Some languages that have emerged have tried to marry the scientific community and information community together in a language that encompasses FORTRAN and provides extensive data handling capabilities. The result has normally been a large statement list accompanied by a set of intricate syntactal rules. The best and most recent example of this kind of effort is PL/1. Most of these languages are said to be more "sophisticated" which is not necessarily desirable. Webster's definition of sophisticated is "deprived of original simplicity; made artificial, or, more narrowly, highly complicated, . . ."

25X1A The guidelines for the System Language development undertaken by [redacted] were simple and few in number. It was felt that the language should be:

1. Oriented to information processing

2. Oriented to the novice and average programmer.

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

Within these guidelines, the language has been created to operate in terms of characters and character strings with the smallest unit of infomation being a single character. Only a moderate amount of arithmetic capability was provided, principally for counters and index control. A study was made to determine the basic functional attributes used in the description and manipulation of information. The language has emerged as a set of 14 statements divided into four functional areas.

Data definition	1
Linkage definition	5
Program definition	6
Input/Output definition	2

In addition to the statement list, six special characters and five notational variations are defined. Five administrative statements have been provided for compiler and listing control. The inclusion or exclusion of each functional statement was based on necessity and utility. The system language, in its current form, represents a simple but powerful tool that can be utilized in the creation of an information system such as the IIS. An individual using the language need have little knowledge of the object computer as the language for the most part is machine independent.

5.6.2.2 Comparison of Alternatives

(1) SPRYE

Advantages

a. No cost to NPIC, either initial or maintenance cost.

b. Simplest language for programmers to learn, (no extensive training is required to use the SPRYE language.

c. SPRYE has the potential of producing the most efficient code for real-time programs, i.e. smaller and faster operational programs.

Disadvantages

a. Assembly language programs have a tendancy to require much more "debug" time than those coded in a Compiler Language.

b. When and if the 494 is replaced by another computer, a major reprogramming effort will be necessary i.e. SPRYE is machine dependent.

c. SPRYE is not the Univac Assembly program for the 494, therefore other programs coded in 494-SPURT (programs obtained from other users of 494's) would have to be modified for NPIC use.

(2) System Language

Advantages

a. The System Language is a tailored information system compiler, therefore is less complex and could be achieved for less cost than a more "universal" compiler, e.g. COBOL, PL/1, or JOVIAL.

b. On the basis of experience with similar languages, it is felt that, with the system language, a 20% reduction in the time between program creation and program completion can be realized. If the system language compiler requires approximately 4 man years to produce, and the overall five year cost of the IIS programming is 258 man years of effort, then a savings of 48 man years can be realized over a five year period.

c. When and if a change over of hardware is accomplished at NPIC, operational programs can more readily (and for less cost) be transferred to a new computer.

Disadvantages

a. The risk of developing a compiler in parallel with operational programs (i.e. operational program are

(3) NSA COBOL and FORTRAN

Advantages

a. No initial cost to NPIC.

b. Maintenance cost could be shared between NSA and NPIC.

c. As in the System Language, a 20% reduction in the time between program creation and program completion could be realized.

d. When and if a change over of hardware is accomplished at NPIC, operational programs can more readily and for less cost be transferred to a new computer.

Disadvantages

a. NSA COBOL and FORTRAN are not yet operational, therefore a risk similar to that discussed under System Language is present.

b. COBOL was designed for batch processing business oriented applications. It's application to real-time information processing is therefore questionable e.g. excessive code generation, inability to segment programs, etc.

5.6.2.3 Conclusion

The major disadvantage to Compilers is one of risk and for Assemblers it is one of cost. It therefore seems that alternative (4) a mix of coding, some programs in assembly language, others in a compiler language, is the most reasonable at this time. By this we mean, the initial IIS operational programs (those implemented in late 1968) should be coded in the SPRYE Assembly Language. When the NSA COBOL and FORTRAN compiler become available, they should be re-examined to determine their utility to future Center operation in comparison to the proposed System Language. At this time of evaluation, the Center should select one or more of the three compilers to replace the SPRYE Assembly Language. This approach will eliminate the risks and capitalize on the cost savings offered by the Compiler approach.

5.7 Cost Benefits to Pre-Readout Activities

This section discusses certain analyses which were performed in an endeavor to ascertain if the costs of preparing the various software programs necessary for Pre-Readout activities could be economically justified.

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

The sections that follow define the operations under consideration, analyze the interrelationships of the operations, and then discuss the products and the costs that might be associated with the programs necessary to produce the products.

5.7.1 Basic Operations Under Analysis

The operations under consideration are those that are currently performed prior to the receipt of film at the Center. In particular we are concerned here with the operations that would utilize the services of the IIS starting with the receipt of the first orbital predictions and camera data. The operations end with the completion of listings and other data internal to the IIS which permits the efficient assignment of film and PI's to Immediate Readout.

For convenience these operations have been broken into three areas, Target Prediction, Target Assignment and Film Can Assignment.

5.7.2 Interrelationships of Operations

The three identified operations are basically tied together in the order that they have been presented. Target Prediction is the cornerstone that supports all activities that can be performed to make early readout activities more effective. These operations are necessary for later efforts in Target Assignment because the knowledge of which targets have been covered determines what the various readout team workloads will be. Target prediction must also precede Film Can Assignment because of the continuous form of the strips of film in the cans. The assignment of the cans must permit the higher priority targets to be read out first and because of the duplication of the film (copies), redistribution can be scheduled to permit efficient utilization of the PI's.

In general terms it is therefore evident that Target Prediction operations in the IIS must be performed in order to provide the data necessary for both Target Assignment and Film Can Assignment. Furthermore, Film Can Assignment is dependent upon data generated as a result of Target Assignment Operations.

5.7.3 Products of Operation

Film Can Assignment

There are three basic products that result from the software programs that would be written to support Film Can Assignment. The first would be the various film distribution lists utilized for First and Second Phase readout. The second would be the film registry list which would permit

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

up to date records to be maintained of the actual status of film. The final product would, in the future, form the basis for an on-line film control system.

Target Assignment

There are three basic products that result from IIS activities contemplated for this operation. These are the team readout assignments, the workload estimates and the distribution information for presentation sheets.

Target Prediction

The IIS will support pre-readout activities with various products such as

Updated Film Coordinate Prediction
NPIC Target Film Coordinate Prediction
On line check list for target reports
Map requirements list
A list of targets in pass order
A list of targets by priority

In addition, internal use of this information is necessary as input into the Target Assignment and Film Can Assignment programs.

5.7.4 Product Cost Analysis

Basic estimates have been made of the programming effort* that will be required to include the operations under discussion in the IIS. It is estimated that 10 man months of programming effort will be required to create the Target Prediction programs for Satellite missions, 4 man months will be required for the Target Assignment programs and 4 man months for the Film Can Assignment programs. It is therefore estimated that 18 man months of effort or approximately [redacted] in programming cost will be required to permit the full range of IIS Pre-Readout Activities if the programming is done by an outside contractor. The costs would vary somewhat if Center personnel were used. 25X1A

A review of information made available together with projections of future manpower estimations made in sections 6.2 and 6.3 indicates that a saving of some 26,000 man hours a year can be effected by the utilization of the IIS for pre-readout activities. If the average cost of a man hour of labor at the Center is more than [redacted] the programming costs would be recovered in less than a year. 25X1A

* It is assumed that a UNIVAC 494 FORTRAN IV Compiler is available at the Center.

25X1

25X1 [redacted] TOP SECRET [redacted] 25X1
Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

If it is desired to include aircraft missions which have INS records it is estimated that the additional programs would cost approximately [redacted] to develop which could be paid for in an additional year or less if manpower costs averaged [redacted] or more per hour.

25X1A
25X1A

It is estimated that the time required to run the program would be insignificant compared to the time required to print out the various lists that would result. Operational costs would therefore be practically negligible. (Printer costs have already been considered in Section 5.5).

5.7.5 Conclusions

It is reasonable to expect that savings in manpower over a two-year period would more than pay for the costs of writing the computer programs necessary to support the activities discussed in this section. It is therefore recommended that pre-readout activities be included in the IIS.

5.8 Data Base Reliability

In section 5.1.3 the problems associated with the catastrophic failure of the data base storage units (i.e., large drums) were considered. It is also necessary to consider that these devices may be operating properly in general but may still provide a certain amount of erroneous data. Therefore consideration has been given to the methods and need to detect errors in the data base including the advantages of sampling for errors, the use of an independent file as a backup for the primary data base, the potential for determining the nature and scope of systematic errors, and programming provisions that should be made for storage hardware maintenance.

5.8.1 Detection Methods

In a data handling system such as the one under consideration there are two kinds of errors that can result in faulty data being delivered at the system output. One is error, or noise, in the data itself and the other is error, or noise, in the control information used to describe and locate the data. The ramifications of detecting each kind of error will be discussed in turn.

5.8.1.1 Errors in the Data

Generally speaking, errors in the data itself are caused by hardware problems and can be detected by having a certain amount of redundancy such as parity bits and check sums included with the data. Furthermore,

25X1
Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3
TOP SECRET [redacted]

experience with computing systems has shown that essentially all of the errors of this type occur in the peripheral equipment (i.e., drums, tapes, card equipment, etc.) as opposed to the main frame of the computer, including its core. Therefore, it appears appropriate to assign responsibility for generating and sensing the necessary redundancy to that portion of the executive program concerned with the input/output operations.

5.8.1.2 Errors in the Control

The situation with respect to errors in the control information is quite different. The errors do not generally arise in the peripheral equipment, but in the main frame itself. They are caused by various program "bugs" and unanticipated combinations of operating conditions. The result is that perfectly logical and internally consistent entries are recorded and/or retrieved from the wrong locations. Generally speaking the program algorithm is being solved correctly but an unanticipated set of inputs results in an unanticipated output. Therefore the error cannot be detected by the conventional kind of redundancy.

One method of detecting such errors is to maintain the same information in two files and compare data retrieved from one with data retrieved from the other. However, to be effective such a procedure should have two different file organizations and two different file update and retrieval procedures. Unless this duality is maintained the same errors will be made with both files.

Another approach to detecting errors is to include labels such as file, unit record, and sector designators as part of the data in each file. These labels are generated along with the data but are not used by the file update routines to determine where to store the data. When the data is retrieved the label is checked against the retrieval parameters to determine whether the proper set of data was obtained. It is felt that this approach will be very effective in detecting control errors that are likely to occur in this system. These include such things as inadvertent overwrite of storage areas, improper revision of index information and improper calculation of physical addresses on both the storage and retrieval cycles. About the only type of error that would not be detected is improper translation from user oriented terms to computer oriented terms, e.g. from a file name to the corresponding file designator.

5.8.2 Requirements for Error Detection

Error detection at the output of the system will certainly catch all of the errors that can be detected, regardless of when in the store-hold-retrieve cycle they might have occurred, and so would be sufficient from the standpoint of minimizing the number of undetected errors. However,

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3
25X1

additional error detection should also be considered to the extent that it eases error correction procedures.

The problems associated with correcting errors that have been detected are a function of when in the store-hold-retrieve cycle the error occurs and when it is detected. For instance, if an error occurs as part of the retrieval process, i.e., the data base and its indices are still correct, a detected error can quite readily be corrected by reinitiating the retrieval request. Similarly, if an error occurs while data is being stored and is detected while the input data is still available it can be corrected by re-entering the data. However, if the input error is not detected immediately or if an error occurs subsequent to the input event (e.g. data or indices are changed inadvertently or it becomes inaccessible because of a partial drum failure) correct data can only be obtained through access to an independent file of the same information.

5.8.2.1 Detection of Errors During Storage

As indicated above, errors made as data are being stored can quite readily be corrected if detected while the input data is still available. A check for errors in the data can be made by requiring the input/output portion of the executive program to read back the data after it is stored and check the parity characters or check sum. This can be achieved quite easily and would detect and allow correction of most of the errors that will occur while data is being stored.

To detect control errors requires that a path be established to the newly stored data by a method that is essentially independent of the one used to store the data. While this could be achieved, using the query procedures for instance, it is not felt that the number of control errors anticipated justifies the computer time required to generate and compare the independent retrievals.

5.8.2.2 Sampling for Errors

Perhaps the most likely cause of erroneous output data is a deterioration of the stored information while it is resident in the storage medium. This can come about through a deterioration of the storage medium itself or through a software error that affects the indices to the data base or allows an inadvertent overwrite of the data. Ideally it would be nice to know of such errors so that they could be corrected before the data is required as a system output. Since there is no clue as to where such errors might lie, the only way to find them is to methodically check the whole data base or sample it randomly. The first is out of the question because of the size of the data base.

25X1
Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

The benefits that could be expected from random sampling are shown on Figure 30 . This chart provides the relationship among the fraction of the data base in error, the number of samples taken, and the confidence, or probability, of finding an error. This shows, for instance, that 10,000 samples would be required to provide a 95% confidence of finding an error caused by a failure such as one of the 64 heads on one of the six storage drums (0.003 of the data base in error). Likewise a sample rate of approximately 3000 samples per hour would be required to provide an 80% probability of finding, within one hour, errors caused by a program problem that destroys the data base at a rate of 0.1% of the data base per hour. These two examples illustrate the large number of samples required to detect the effects of significant failures. Basically, it is the very large size of the data base that causes the required number of samples to be so high. For these reasons it does not appear worthwhile to continuously sample the data base looking for errors. The queries that are processed as part of the normal system activity will serve this function as far as locating sources of errors are concerned.

5.8.2.3 Conclusions

All data sets retrieved should be checked for errors in the data and in the control information by the methods described in Section 5.8.1. If an error is found the retrieval should be reinitiated at least once to see if an error free set can be obtained. If the error persists, recourse must be made to an independent source of data as described in 5.8.3.

In addition to checking retrieved data, the system should also check for errors that might be generated in the data during the storage cycle by reading back newly stored data. Errors introduced into the control information during the storage cycle will be detected during the retrieval cycle.

5.8.3 Independent File

As indicated above, provision should be made for an independent source of data that can be used to correct errors detected in the primary data base.

5.8.3.1 File Alternatives

One form of backup that might be considered is to maintain a duplicate file on duplicate drums. While this would provide backup for breakdowns in the drum itself, it would not provide backup for breakdowns in the drum itself, it would not provide backup for breakdowns in the software controlling the update of the drums.

25X1
Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

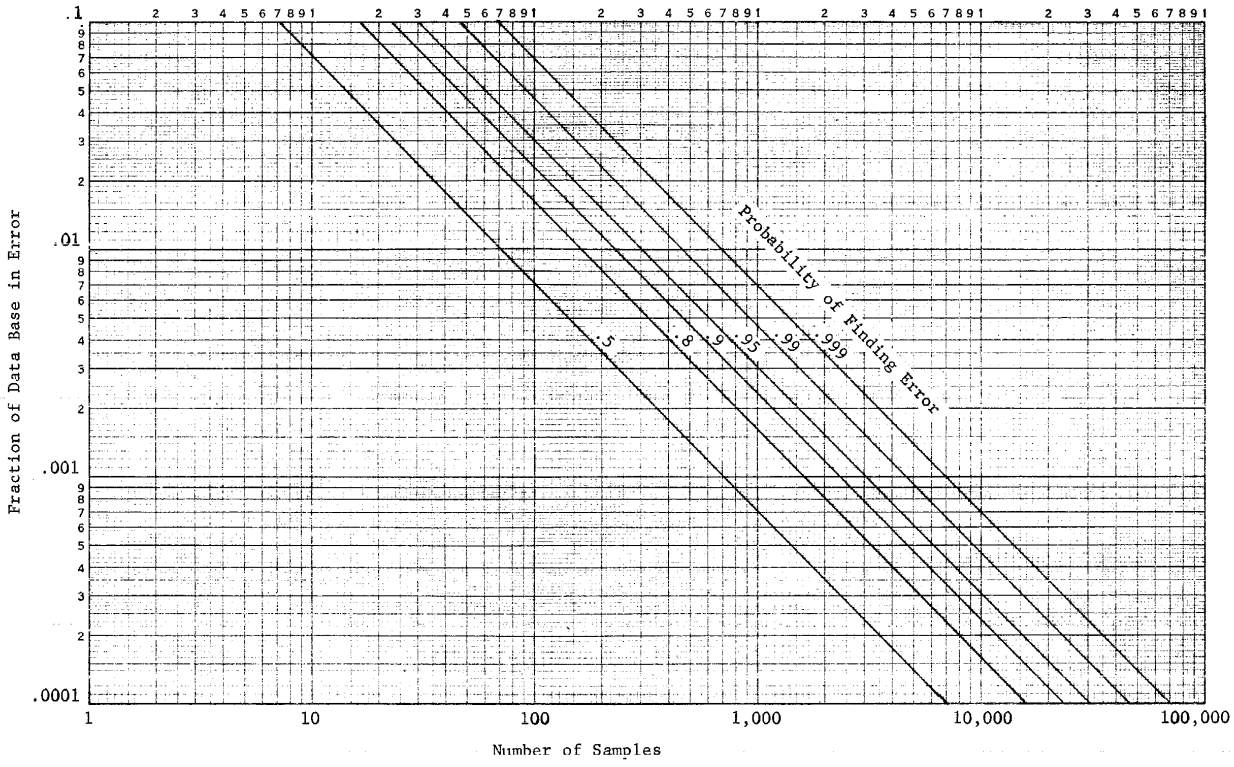


Figure 30 Effectiveness of Sampling

TOP SECRET

TOP SECRET

25X1

25X1

SCR 352

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

Another method is to periodically make a copy of the primary storage medium, checking it as it is made. In addition the inputs to the data base update program are also recorded. The combination of the two allows the latest version of any portion of the data base to be reconstructed. While this method is more cumbersome to use for making corrections than the other one described above, it provides more extensive backup and therefore is the one recommended.

5.8.3.2 File Reliability

It should be noted that the purpose of the independent file is only to correct errors detected in the primary file. It is not used as a base for comparisons to determine the presence of errors. As a result the reliability requirements on the independent file are quite different than they would otherwise be. For instance, if the primary file has a detected rate of one erroneous data set in 10^4 and the rate of the independent file is as high as one in 10^2, the effective, combined rate becomes one in 10^6. Furthermore, if, for the detected error rates above, the undetected error rates are one in 10^8 and one in 10^4 respectively for the primary and independent files, the effective undetected error rate is only degraded to two in 10^8. Thus it can be seen that even a relatively unreliable method of storing the independent file can make a significant improvement in the overall system provided that it is only used to correct errors detected in the primary file and that it has error detection provisions.

While it is of course desirable that the system error probability or error rate be made small, it is not necessary that it be reduced to zero. The primary material from which the file data was originally developed is available and can be consulted if necessary.

5.8.4 Analysis of Errors

The error detection providisions recommended in Section 5.8.2.3 will, in the course of the system operation, indicate that certain data is in error. It would seem worthwhile to determine whether the errors detected are simply random errors or are symptomatic of some fault in the system. This can be done by providing programming that investigates locations in the data base that are physically or logically adjacent to the detected error. The investigations should check for errors in data and errors in control as appropriate. In addition it will in some cases be useful to investigate data base updates that are adjacent in time to the most recent update of the location containing the detected error. The latter capability

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

With proper analysis the location and extent of any systematic errors can be determined and corrective action initiated. This can include operator notification of the fault and the regeneration of the affected data from the independent file and the record of data base updates.

5.8.5 Storage Hardware Maintenance

Since the IIS is envisioned as a system that eventually operates 24 hours a day, seven days a week, it follows that it will be desirable to perform diagnostics and maintenance on portions of the storage drums without completely removing the drum from system operation. The programs proposed in Section 5.8.4 for investigating the data in various storage locations should be made available to service engineers. In addition the storage allocation programs should have provision for removing data from designated portions of storage so that those portions are available to the service engineers for special diagnostics and maintenance. The removed data would be placed in another portion of storage so as to be available to the system.

5.9 UNIVAC 490/494 Software Survey

Because of the possibility that some of the more common features of the IIS (retrieval, editing, etc.) might presently exist as working software in other 490/494 installations, a limited survey of UNIVAC 490/494 users was conducted to determine the existence, suitability, and availability of software which might serve the needs of the IIS. The level of effort expended for the survey was not large, but provided the desired assurance that potential savings in IIS software implementation were not overlooked. The approach taken in the survey was (1) identify all UNIVAC 490/494 installations, (2) determine which installations would potentially have applicable software, (3) contact this list of users to determine suitability and availability of software to serve the needs of the IIS.

5.9.1 UNIVAC 490 Series Installations

The following is a list of 490 Series Installations as of September, 1967:

25X1A

25X1A [redacted]

5.9.2 Installations With Similar Applications

The installations marked with an asterisk in the above list are those which were deemed most likely to use software applicable to the IIS. It should be noted here, [redacted] 25X1A

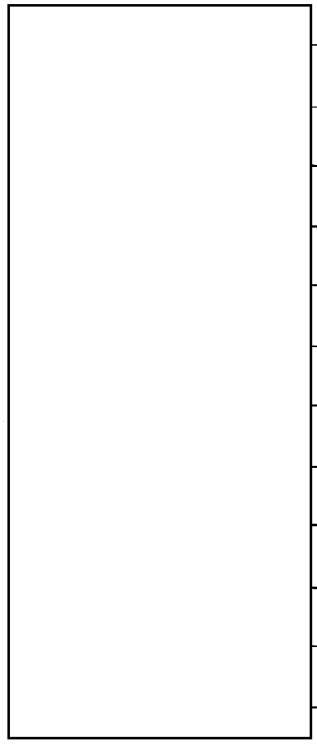
25X1A [redacted]

RUM	- 494 Executive System
ART	- Analytic Real-Time Program
TIPS	- Technical Information Processing System
SPRYE	- 494 Assembly Program
COBOL	- Common Oriented Business Language
FORTRAN	- Formula Translator
Utility Library	- Includes debugging aids

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

25X1A



COMPUTER CONFIGURATION	MAJOR COMPUTING PURPOSE	COMMENTS
2 494's (65K)	INVENTORY ANALYSIS	
1 490 (32K) 2 FASTRANDS	JOB PLANNING	USE SPURT FD, COBOL 61, AND FORTRAN II
2 492's (65K) 9 FASTRANDS	SALES SUPPORT MANAGEMENT CONTROL	
2 494's (65K)	AIRLINE RESERVATIONS	HAVE OWN EXEC. SYSTEM- CONTORTS 100% SPURT CODED
1 490 (32K) 2 FASTRANDS	PRODUCTION CONTROL	HAVE USED COBOL 61 BUT NOT SATISFIED WITH IT
2 494's (65K) 2 FASTRANDS	CUSTOMER ACCOUNTING	
2 494's (65K)	CASUALTY PROPERTY AND AUTO POLICIES COMPUTATION	
2 490's (32K) 11 FASTRANDS	NAVAL LOGISTICS	
1 491 (65K) 1 FASTRAND	INVENTORY CONTROL AND BILLING	
1 490	ANTENNA POINTING	NO PERIPHERALS - VERY SPECIALIZED SOFTWARE
2 491's	MOTEL RESERVATION SYSTEM	

25X1

Figure 31 - 490 SERIES USER SURVEY

TOP SECRET

25X1

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

all users who were contacted are not listed in Figure 31 for one or more of the following reasons, (1) the contact felt his operational programs were proprietary, (2) lack of interest in the survey, or (3) contact agreed to forward information but has not followed through.

It was generally found that operational software is unique to a given installation. This is the case because of the non-uniformity of input/outputs and file structures of the installations contacted. The only applicable available software identified by this survey falls into the "utility" or support area, i.e. compilers, assemblers, trace routines, executives, etc. These utility programs are more universal over a given class of computer and could be used in implementing the IIS. However, all software of this type is already in use at the Center (or available from NSA).

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

6. COST/EFFECTIVENESS EVALUATION

6.1 Measures of Cost/Effectiveness

A set of design and evaluation criteria were developed as part of Phase I of this program to form a basis for judging the relative worth of the various alternative concepts that were considered. These criteria were suitable for use at the conceptual stage, but are not directly applicable, in all aspects, to the comparison of two specific designs; i.e., the IIS and an extrapolation of the current system. Many of the criteria were defined only in subjective terms. In addition, the several criteria were related through weighting factors that were established subjectively. Thus, a system rating based on these criteria is highly dependent on the judgment that is part of these subjective factors. This was satisfactory when considering alternative concepts for the IIS since details were not available for objective measures and similar judgments could be applied in all cases considered. However, it is not particularly satisfactory when comparing two specific designs that differ widely in many aspects. Therefore it has been necessary to develop another method of measuring the relative worth of the two systems which is nevertheless based on the factors used in the original criteria.

One goal in the development of the method has been to eliminate the necessity to quantify or measure parameters that are essentially subjective. This is not to say that they should not be considered; it is only to say that they should not be included in a numeric measure of the system worth, but rather they should be considered in parallel with the numeric measure.

The second goal has been to reduce the number of subjective weighting factors required and thus reduce the sensitivity of the measure of worth to judgments. This has been done in part by reducing the number of individual factors to be included in the measure of worth without significantly reducing its comprehensiveness. This can be done now that specific designs are being considered (e.g., since the system cannot fail to produce the necessary outputs, reliability need not be considered separately but only to the extent that the necessary backup procedures effect system costs and timeliness of system outputs.) As another part of achieving this goal, the measure of worth has been taken as the cost/effectiveness ratio, rather than a sum of terms as was done in Phase I. This eliminates the need to provide a weighting factor between the cost factors and the effectiveness factors, a difficult task at best.

6.1.1 Cost Factors

Of the ten criteria used in Phase I, four can be considered as cost factors, i.e. part of the "price that must be paid" to realize the benefits of the system. These were facility requirements, personnel requirements,

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3
25X1

total costs, and phase-in. Since establishing these criteria, it has been determined that the two systems do not differ significantly in the size or criticality of facility requirements. Further, any special requirements of one system over the other can be adequately accounted for in system costs. Similarly, no significant phase-in problems that can be treated objectively have been identified other than those associated with cost, e.g. reformatting the data base. Therefore, these two factors will only be considered to the extent that they affect system costs.

Personnel requirements should be considered from two standpoints. First the use of personnel represents a cost equivalent to their salary and some allowance for the cost of fringe benefits, office space, supervision, etc. Second, it is worthwhile to reduce the personnel requirements since Tables of Organization are usually fixed independently of the budget. In addition, there may be a consideration of certain critical skills that are difficult to obtain. Therefore, while personnel requirements are treated in terms of the associated costs, provision is made to weight this cost in accordance with a subjective judgment of the additional importance of personnel over and above the associated cost. Furthermore, provision is made for considering two categories of personnel, i.e., PI's and support personnel.

The determination of the costs that should be included depends on what is being measured, the cost effectiveness of the IIS or the cost effectiveness of the activities being supported. If it is the former, only the directly associated costs should be considered, but it is difficult to establish a clear definition of costs to be included in the extrapolation of the current system. If it is the latter, all costs associated with the supported activities should be considered. Since the IIS is considered as a tool to improve the operations of the Center, the latter position is basically the one taken.

For the purposes of determining cost effectiveness of a continuing operation, it is appropriate to consider costs on a per unit time (e.g. annual) basis. Personnel and equipment rental costs are certainly of this type. The price of purchased equipment can reasonably be converted to a monthly cost by dividing purchase price by 40. This makes an allowance for maintenance and spare parts. In the case of the IIS it is also reasonable to prorate one-time costs such as data base conversion and programming over a five year period.

6.1.2 Effectiveness Factors

The other six of the ten criteria used in Phase I were stated in terms of effectiveness factors, i.e. as measures of how well the system contributes to the carrying out of the role of the Center. These were

25X1
Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

performance, form and organization of material, reliability, flexibility, expandability, and compatibility. The applicability of each to the current evaluation effort is discussed below.

6.1.2.1 Performance

Performance is generally taken to be a measure of such attributes as accuracy, completeness, throughput, and timeliness. It was established in Phase I that accuracy is not an appropriate criterion in the evaluation since the system is not allowed to degrade the information it handles and is not able to improve upon the results of the intellectual process of the PI. The system can effect completeness by providing the PI with the proper support materials. However, the rating of this factor is essentially subjective and so it has not been included in the objective measure of effectiveness. The throughput of the system is specified as part of the input/output loading factors (e.g. numbers of missions, targets, reports, etc.) and, assuming these requirements are met, is not subject to evaluation.

Only timeliness was considered in the performance factor during Phase I. It was described in terms of "age of information" and "response to a request." To provide a more definitive measure of timeliness the three basic activities of the Center (i.e., immediate readout, detailed reporting, and data base operation) have been investigated to determine the elements which are critical from a time standpoint.

There are four times that could be considered in the immediate readout activities: the time to prepare for a mission, the time to produce an Oak Report, the time to produce the Oak Supplement, and the time to complete Second Phase readout. Since generally the same techniques are used to support the latter three events, a measure of any one should adequately reflect the effectiveness of the system in these areas. While time to prepare for a mission may become important in the future, it is not considered critical at present and has not been included in the measure of effectiveness in the interests of keeping the number of factors small. Therefore, the timeliness measure taken from immediate readout activities is the time to produce one of the reports. The one chosen is the Oak Report and the measure is defined as the elapsed time between receipt of the film and the completion of the mats that are used to print the report. The number of targets to be included in the report is the number covered by a typical satellite mission as described by the input/output loading factors.

In the area of detailed reporting, time by itself was not taken to be a factor of primary importance. These reports will generally be produced on a scheduled basis rather than in response to a request that needs

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

a rapid response. The time to produce them is important from the standpoint of the manpower consumed, but this is considered to the extent possible in the cost factor. While there are reports prepared in response to requests, they are so varied in nature as to defy the definition of a typical "response time." Therefore, no time factor for detailed reporting has been included in the effectiveness measure.

There are two aspects of the data base operation that can be considered. One relates to the "age of information" and is measured by the time required to add information to the data base. Information to be added includes Center readouts, field reports, statements of readout requirements, exploitation products data, etc. The measure is the elapsed time between the approval of information for entry and its availability for queries. The other is the time to respond to a query. Because of the importance of the data base to the Intelligence Community, both of these times have been included in the effectiveness measure, although it may be desirable to weight them less heavily than the time to produce an Oak Report.

6.1.2.2 Form and Organization

This factor is important from two aspects. First, the individuals using the system outputs will be able to do their jobs faster if the information is well organized and presented in the proper form. Second, with the proper form and organization, these individuals will be less likely to overlook pertinent information and thus will be able to perform their tasks with greater proficiency. Without extensive testing, the second aspect is purely subjective and so is not included in the measure of effectiveness. While the first aspect would also require testing for an exact measure, it is felt that reasonable time savings estimates can be made. However, this saving will be accounted for in the time factors under performance and as part of the manpower requirements in costs and so need not be considered separately.

6.1.2.3 Reliability

This factor was meant to be a measure of the vulnerability of the system to component failure and the degree of backup necessary to produce system products in the event of failure. As indicated in the analysis of product preparation (Section 5.3.2), the effects of reliability can be adequately reflected in the time to produce system products and the costs of the necessary backup procedures. Therefore, each time factor will include the average delay that can be caused by component failures and a separate measure of reliability is not required.

25X1

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

6.1.2.4 Flexibility

While flexibility has some importance from the standpoint of the elapsed time required to implement a change, perhaps its primary importance is from the standpoint of the manpower required to implement the change. To measure the former requires the definition of a specific change that is typical of those of concern. Therefore, this factor is measured in terms of the manpower required to effect the changes in mission parameter formats, report formats, etc., that can normally be expected over the course of a year. This manpower is included in the cost factor.

6.1.2.5 Expandability

Expandability was defined as a measure of how readily the capability of the system could be increased to carry out either greater workloads or new tasks. While potential expansions could be defined, the definitions would be arbitrary and the importance of accommodating them highly subjective. Therefore, this factor has not been included in the measure of effectiveness.

6.1.2.6 Compatibility

This criterion was intended to measure how well the IIS supports automated and non-automated functions outside the IIS (e.g. COINS). Since developing the concept, the support of the functions considered has been taken as a requirement and there are no objective measures of "how well" other than the times described in Section 6.1.2.1. Therefore, no separate factor for this criterion is included in the measure of effectiveness.

6.1.3 Composite Measure

A useful method of judging alternative systems is to compare the ratios of their costs to their effectivenesses. These ratios are in essence the "cost per unit effectiveness" and presumably the lower the ratio the better the system. This is not meant to imply that this ratio should be the only consideration. Certainly there is a lower bound on effectiveness below which the system would be worthless. At the other extreme there is an upper bound on cost, above which the system would not be bought, regardless of the effectiveness achieved. Care must also be taken that the factor or factors put in the denominator truly represent effectiveness to the operation being considered. Nevertheless the cost effectiveness ratio has proved useful in comparing alternative methods.

The specific measures of effectiveness that were identified in Section 6.1.2 were all time factors. The concern now is to relate these

25X1
Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3
TOP SECRET

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3
25X1

times to effectiveness and to combine them into a composite measure of effectiveness. While it is generally true that the shorter the time, the greater the effectiveness, it does not necessarily follow that effectiveness is inversely proportional to time. For example, the effectiveness of responding to COINS is not doubled if the time to respond is reduced from twenty milliseconds to ten milliseconds because the person making the query could not tell the difference. However, in the cost performance tradeoffs that were made as part of the design studies the performance chosen is such that these extremes are avoided and the effectiveness in each performance area can be considered as being inversely proportional to time.

The various time factors can be combined either by addition or multiplication. Simple addition is not considered appropriate since the largest term would dominate the sum. Even if the terms were normalized and then added, the term with the greatest difference from the normal value would dominate and improvements in the others would hardly be noticed. A straight multiplication of the three time-terms would be misleading in that halving each time would increase the cost effectiveness ratio by a factor of eight. However, if the geometric mean of the terms is used, this problem is avoided. This in effect is

$$\text{Effectiveness} = \sqrt[3]{\frac{1}{T_1} \times \frac{1}{T_2} \times \frac{1}{T_3}} = \frac{1}{\sqrt[3]{T_1 T_2 T_3}}$$

Furthermore, the times can be given relative weights by considering the form:

$$\text{Effectiveness} = \frac{1}{(T_1)^a (T_2)^b (T_3)^c}$$

where $a + b + c = 1$

The cost factors identified were the personnel and other costs. It is reasonable to sum these factors since the primary penalty of personnel is their costs. However, personnel costs can be weighted if personnel have a significance apart from costs. This gives the following relationship for the cost effectiveness ratio:

$$\frac{\text{Cost}}{\text{Effectiveness}} = (C_o + sC_s + pC_p)(T_o)^a (T_u)^b (T_q)^c$$

25X1
Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

Where: C_s = cost of support personnel

s = weight of support personnel costs with respect to other costs

C_p = cost of PI's

p = weight of PI cost with respect to other costs

C_o = other costs

T_o = time to produce Oak Report

a = relative weight of Oak Report time

T_u = time to update data base

b = relative weight of data base update time

T_q = time to respond to a query

c = relative weight of query response time

$a + b + c = 1$

This is the measure of system worth that will be used in the evaluation. Initially the weighting factors s and p will be taken as unity and a, b, and c as 1/3 each. When specific values C_o, C_s, C_p, T_o, T_u, and T_q are determined in Sections 6.2 and 6.3 below, the effect of these weighting factors on the cost effectiveness ratio and the relative effects on the ratio for two systems can be investigated.

6.2 Extrapolation of Current System

This section describes a method by which the currently projected central processor equipment could be extrapolated into the time period of the IIS. Furthermore, the additional personnel needed for this time period are also estimated.

6.2.1 Personnel

The actual manpower needs of the system will increase over present day levels because of two factors, e.g., the increase in the number of missions and the increase in the number of targets. A total of 48

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3
25X1

activities within the Center were selected as representative of the activities which are closely allied to the proposed IIS. This list is shown as Table XVII. The list was broken into the 5 main activities shown on Table XVIII and considerations were given to the increased workloads in each major activity brought about by both increased numbers of missions and increased numbers of targets. As a benchmark, data pertaining to FY 1967 missions and targets were used. The Extrapolation Figure shown in Table XVIII is, of course, the product of the two factors.

Table XIX summarized the results of using manpower figures for FY 1967 provided by the Center as the benchmark for the 48 identified activities and multiplying them by the extrapolation factors. It can be seen that the average manpower increase for all the activities is 81%. The estimated increase for PI's is 83% and for other personnel is 77%.

The other category of personnel that should be considered in the extrapolated system are the numbers of IPD programming and IPD operational personnel directly supporting those operations which will eventually be a part of the IIS. It is estimated that at the present time approximately 55 persons are in this direct support category. This group will have to grow at a rate of approximately 5 persons per year to keep up with the demands of the rest of the Center. It is therefore estimated that in the time frame of the extrapolated system an average of approximately 80 persons will be needed.

6.2.2 Equipment Costs

It is very difficult to clearly estimate the new equipment that would be utilized in an extrapolation of the current system because of the dynamic growth of the computer field. Discussions with Center personnel have, however, pointed out the current thinking along these lines and it is reasonable to assume that because of community pressures to implement COINS, a substantial portion of the data base would be put on-line.

The assumption has been made that approximately 1/3 of the data base would be available on-line and previous analysis (see Section 5.1.2) has shown that approximately 98.5% of all data requests could be answered with this data. The remainder of the data base is assumed to be maintained on magnetic tapes and in hard copy form. The on-line data base would be largely implemented with approximately 14 dedicated magnetic tape drives and 6 additional tape drives for hard copy and other processing. These tape drives would be serviced by two tape control units.

In addition it is assumed that two FASTRAND Drums would be included in the system. These drums would hold both computer programs and indexes to the data on tape. In addition, mensuration parameters for missions of current interest would be temporarily stored on the drums.

25X1
Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

TABLE XVII

List of Representative Activities

1. Mission Readiness
   - 1.1 Manipulation of Phase I Predicted Targets
   - 1.2 Establishment of Teams and Team Assignments
   - 1.3 Printout of Blip Sheets - Phase I
   - 1.4 Manipulation of Target Folders
   - 1.5 Initial Coverage Plots (1:9M)
   - 1.6 Labeling, Distribution and Tracing of Film Cans
   - 1.7 Phase II Target Prediction
   - 1.8 Printout of Blip Sheets - Phase II
2. Readout Phase I
   - 2.1 COMIREX 1A and 1B Targets
     - 2.1.1 Study Support Material
     - 2.1.2 Location of Film
     - 2.1.3 Readout
     - 2.1.4 Preparation of Work Sheets
     - 2.1.5 Generate Other Report Material
     - 2.1.6 Keypunching
     - 2.1.7 Verify, Edit and Approve
     - 2.1.8 Printout "Draft" Copies
     - 2.1.9 Prepare Final Mats
     - 2.1.10 New Target Acquisition
   - 2.2 COMIREX IC Targets
     - 2.2.1 Study of Support Material
     - 2.2.2 Location of Film
     - 2.2.3 Readout
     - 2.2.4 Preparation of Work Sheets
     - 2.2.5 Generate Other Report Material
     - 2.2.6 Keypunching
     - 2.2.7 Verify, Edit and Approve
     - 2.2.8 Printout "Draft" Copies
     - 2.2.9 Prepare Final Mats
     - 2.2.10 New Target Acquisition
3. Readout Phase II
   - 3.1 Study of Support Material
   - 3.2 Location of Film
   - 3.3 Readout
   - 3.4 Preparation of Work Sheets
   - 3.5 Generate Other Report Material
   - 3.6 Keypunching
   - 3.7 Verify, Edit and Approve
   - 3.8 Printout "Draft" Copies

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

3.9 Prepare Final Mats
3.10 New Target Acquisition

4. Detailed Reports
   - 4.1 Identification of Pertinent Material
   - 4.2 Location of Pertinent Material
   - 4.3 Readout
   - 4.4 Preparation of Initial Draft
   - 4.5 Typing and Proof Reading
   - 4.6 Editing
   - 4.7 Preparation of Final Mats

5. Other Activities
   - 5.1 Data Base Update
   - 5.2 Mensuration Parameters Data Preparation
   - 5.3 Preparation of Briefing Board Notes

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

TABLE XVIII

MANPOWER EXTRAPOLATION FACTORS

		Factor Due to Mission Increases	Factor Due to Target Increases	Extrapolation Figure
1.	Mission Readiness	1.16	1.71	1.99
2.	Readout Phase I			
	COMIREX 1A & 1B Targets	1.16	1.20	1.39
	COMIREX 1C Targets	1.16	1.20	1.39
3.	Readout Phase II	1.16	1.87	2.18
4.	Detailed Reporting	1.00	2.00	2.00
5.	Other Activities			
	Data Base Update	1.16	1.42	1.65
	Mensuration Parameters Data Preparation	1.16	1.00	1.16
	Briefing Board Notes	1.16	1.20	1.39

TABLE XIX

MANPOWER SUMMARY AND EXTRAPOLATION

	FY 1967 Manhours	Extrapolated Manhours
Photo Interpreters	207,650	380,000
Support Activities	124,051	221,000
Total	331,701	601,000

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

The major equipment associated with the central processor and storage devices are shown along with the purchase and maintenance costs in Table XX.

6.2.3 Performance of the Extrapolated System

The basic extrapolated system is an on-line storage and retrieval system centered around a dedicated magnetic tape drive storage medium. The extrapolated system has been expanded into a system of this type primarily to support national and community requirements. Provisions have, therefore, been made available in the extrapolated system to permit the establishment of the equivalent of the IDF, MPF, EPDF, and ODF in support of the National Data Base concept. In addition, a large portion of the active parts of these files have been placed on-line in order to effectively permit COINS queries to be answered within a reasonable time span.

The remainder of the operational characteristics of the extrapolated system are quite similar to current operations. It is assumed that basic data base updates will be performed periodically in a batch processing mode. It is assumed that the basic generation and distribution of Target Briefs will be accomplished also as a batch printout and manual sort operation. The current on-line mensuration activities will also be continued.

In view of the system just described, it is estimated that the average time before newly approved information will actually be placed in the data base will be approximately 4 first shift hours. This is based upon the assumption that data base updates will be accomplished each day during second or third shift and that data is generated randomly during first shift activities.

It is furthermore assumed that the average time required to answer a query will be approximately 10 minutes. This assumption is based upon the realization that the most active part of the data base will be available on-line and that the major query load will be generated by COINS.

The average time to prepare the First Phase readout on high priority targets is based on an extrapolation of current performance. In FY67, the average time for the 24 missions processed was approximately 2.5 days. Analysis of data from the MIS indicates that on the average, approximately 100 PI's were involved in this readout. Since the number of high priority targets is expected to increase by approximately 20% it can be anticipated that the average time would increase to about 3 days if the number of PI's is held constant.

The above estimates are considered to be reasonable based on current Center practice.

25X1

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

TABLE XX

SUMMARY OF EXTRAPOLATED SYSTEM EQUIPMENT COSTS

25X1A

Equipment	# Required	
Central Processor		
UNIVAC 494	2	
FH 880 Drum	1	
FH 880 DCU	1	
Com. Terminal	1	
CTM Medium Speed	1	
CTM Low Speed	1	
Miscellaneous	As required	
Sub Total		
Storage Devices		
Magnetic Tape Unit	20	
Magnetic Tape Controls	2	
Fastrand II	2	
Fastrand II Control Unit	1	
Sub Total		
TOTAL		

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3
25X1

6.3 IIS Cost and Performance

This section provides the cost, personnel, and performance factors required for the cost effectiveness evaluation. The development of the factors is based on the system description and analyses given in the preceeding sections.

6.3.1 IIS Personnel

The basic estimates prepared in Section 6.2.1 were modified and adjusted in estimating the manpower requirements of the IIS. Of the 48 previously identified activity areas it was found that manpower savings could be realized in 28 activities. The activities were mainly in the support area and it was assumed that the effect of the IIS upon the Photo Interpreters will be not only increased quality and scope of reporting but also a 20% reduction of time. Table XXI has been prepared which lists the estimated man-hour requirements in the 5 major activities under FY 1967, the Extrapolated System and the IIS System.

Table I illustrates the fact that the personnel savings are in all areas, i.e., support personnel, Photo Interpreters, and, of course, relatively large savings in the number of IPD operational personnel needed after the IIS is operational.

6.3.2 IIS Equipment Costs

A summary of the IIS Equipment Costs has been included as Table XXII. This table lists the major equipment associated with the Central Processors, the storage devices and the consoles along with their purchase and maintenance costs.

As was pointed out in Section 5.1.3.3 approximately 2/3 of the data base will be maintained on 4 of the FASTRAND II Drums with the remaining 3 utilized as backup devices. Approximately 99.9% of all queries will be answerable from this on-line data base. The remainder of the data will be stored on magnetic tapes and in hardcopy form.

The 10 magnetic tape units are included for other processing tasks and can be utilized as a second level backup to the drum system.

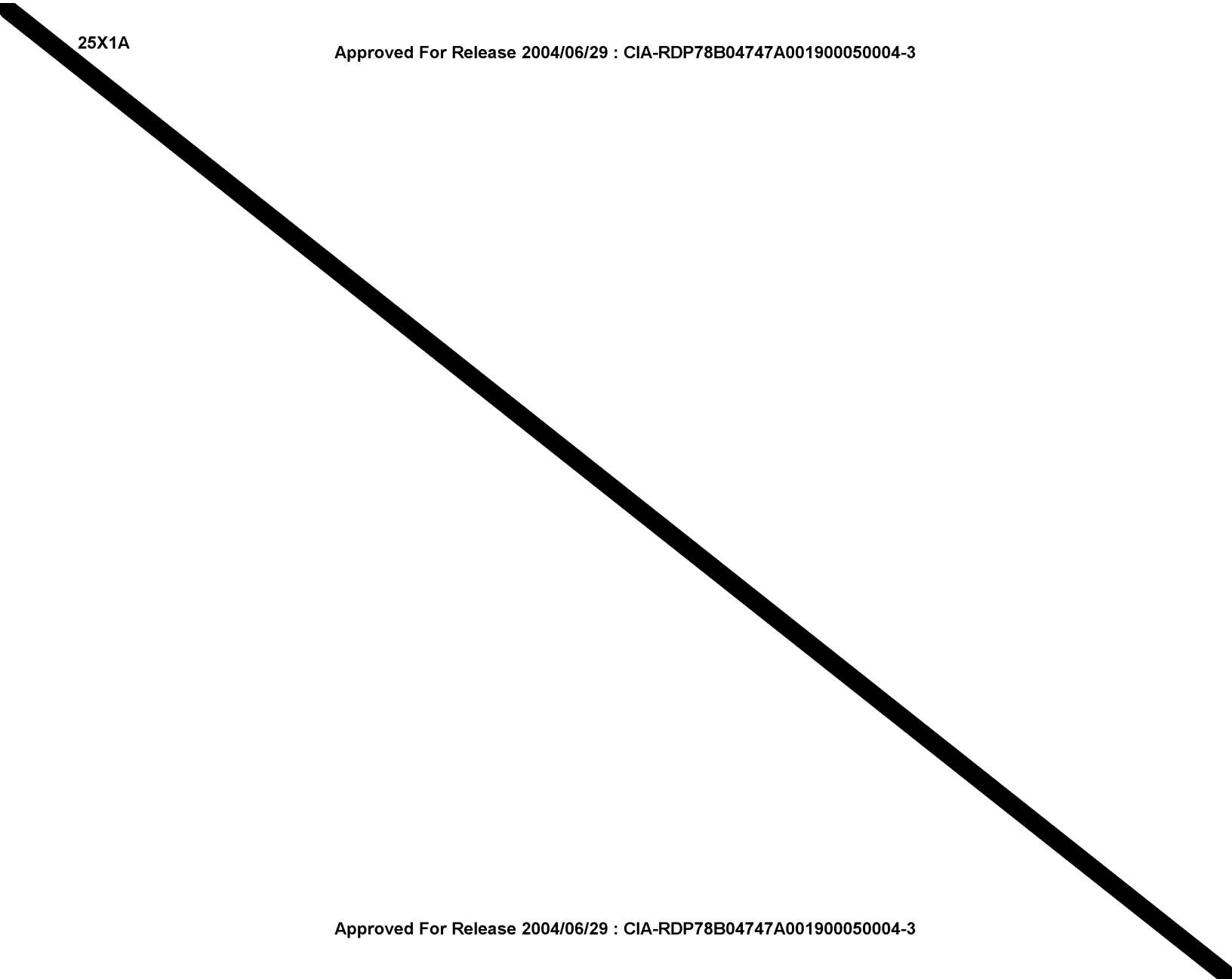
6.3.3 Performance of the IIS System

The basic aspect of the IIS that is considered in this section is an on-line storage and retrieval system centered around a data base which is maintained on dedicated magnetic drums. The basic system has a first level backup provided by additional drums and an additional backup of magnetic tapes and hard copy. Basic entry into the data base is by

25X1
Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1A

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3



Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

TOP SECRET

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

TABLE XXII

SUMMARY OF IIS EQUIPMENT COSTS

Equipment	# Required
UNIVAC 494	2
FH 1782 Drums	2
FH 1782 DCU	1
Com Terminal	3
CTM High Speed	15
CTM Low Speed	20
Miscellaneous	As required
Sub Total	
Storage Devices	
Magnetic Tape Units	10
Magnetic Tape Controls	1
Fastrand II	7
Fastrand II Control Unit	2
Sub Total	
Consoles	
SUSIE	7
Edit, U/L Case	23
Data Entry U/L Case	7
TTY KBD/Print 25 Char./Sec.	10
* TTY Print Only 25 Char./Sec.	14
TTY KBD/Print/ Punch 25 Char./Sec.	1

25X1A

random access and analysis of the load and queues has determined that even under "worst hour" loading conditions wait time for service will be sufficiently short to permit a dialog type of access. The analysis has taken into consideration not only the loads associated with queries but also those loads which are generated as a result of data base maintenance and file updating. The analysis indicates that:

a. Many of the operations of the Center will be directly supported by the IIS but will basically require the same amount of clock time to perform.

b. The basic operations involved in reading out a target will not be greatly changed because of the IIS but it is believed that the much more accessible data base will result in a higher caliber of professional work being done by the Center Photo Interpreters in the somewhat shorter times previously mentioned.

Insofar as the comparison of the IIS with the Extrapolated System is concerned only with response times in three representative areas, only these areas will be considered below.

a. As has been pointed out, the actual update of the data base is a continuing on-line process. Since no appreciable queues are anticipated and because it is estimated that individual data element locations can be located in approximately 250 milliseconds, it is reasonable to believe that a basic data base update can be accomplished in considerably less than 1 minute. Considering that several indexes to the data may also have to be updated each time a change is made it is reasonable to allow a factor of 5 to permit for not only updating the data but also updating all means of accessing the data. Accordingly, a conservative estimate of data access time is less than 5 minutes.

b. The analysis of the usage of data has indicated that a query of the data base will require on the average 5 accesses to storage locations. This can be accomplished in less than 2 seconds even under worst hour load conditions. Because of the insensitivity of the effectiveness measure in this region of time, the average query response time has been arbitrarily increased to 6 seconds for evaluation purposes.

c. Analysis of data provided by the Center and certain assumptions about relative numbers of high priority targets indicate that during FY 1967 approximately 5 hours of PI readout time per high priority target was achieved. Insofar as the IIS expects to effect a 20% saving in PI readout time (because of better scheduling, condensed presentations and worksheets), it is estimated that 400 high priority targets can be read out by 100 PI's in about 16 hours distributed over 2 working days. Using an estimate of 1.5 hours to accomplish film delivery on the morning of the first day and an estimate of 55 minutes to enter, verify, edit and

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

approve the last readout, edit the final copy and run the mats on the second day permits a basic 2-day cycle for producing the reports of the readout of the high priority targets. It is, therefore, estimated that completion of the Oak readout under the IIS should be considered to require 2 days.

6.4 Cost Effectiveness Ratios

The factors for the cost effectiveness ratios for the two systems are:

	Extrapolated System	IIS
Cost of Support Personnel		
Cost of PI's		
Other Costs		
Time to produce Oak Report		
Time to update data base		
Time to respond to a query		

25X1A

Using these values in the cost effectiveness relationship developed in Section 6.1.3 with unity weights for cost factors and 1/3 for each time factor gives a cost effectiveness ratio for the extrapolated system of 455×10^6 and for the IIS of 24×10^6. This indicates that the cost per unit effectiveness for the extrapolated current system is 19 times that for the IIS.

The large advantage of the IIS is a result of the very much smaller time required to update the data base and respond to queries and the fact that they are weighted to be two-thirds of the effectiveness. The effect of assigning other weights to these factors is shown on Figure 32. This is a plot of the ratio of the cost effectiveness ratios of the two systems as the weights on the three time factors are varied in such a way that their sum is unity. The asterisk is the point where each weighting factor is one-third. The dashed line is the limiting condition where the weighting factor on the Oak time is zero. This indicates that the IIS is always better than the extrapolated system as long as the weights on the time factors are positive.

A similar plot is given on Figure 33 showing the effect of varying

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

Relative Weight of Oak Time = 1-b-c

$$r = \frac{\text{C/E for Extrapolated System}}{\text{C/E for IIS}}$$

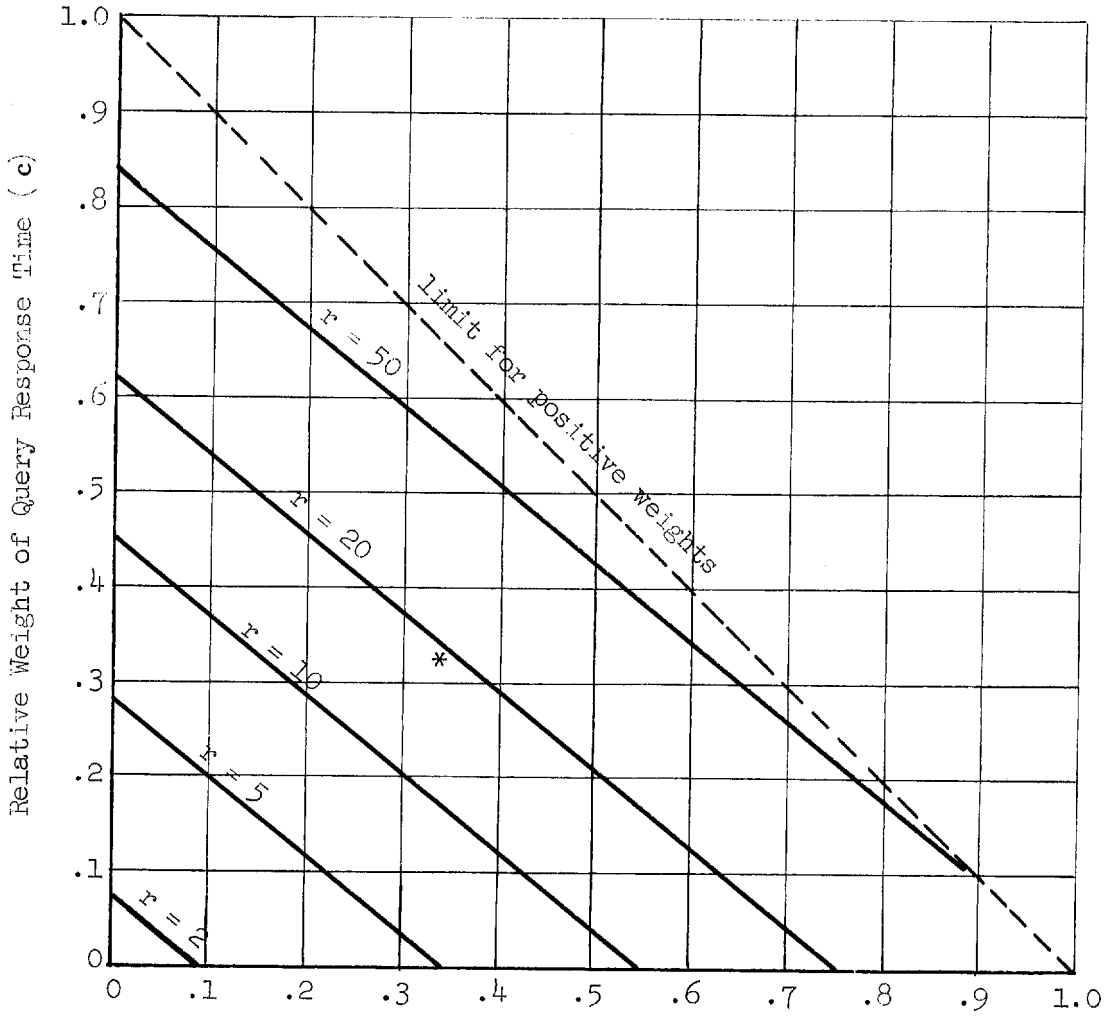


Figure 32 Effect of Varying Weights of Time Factors

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

$$r = \frac{\text{C/E for Extrapolated System}}{\text{C/E for ILS}}$$

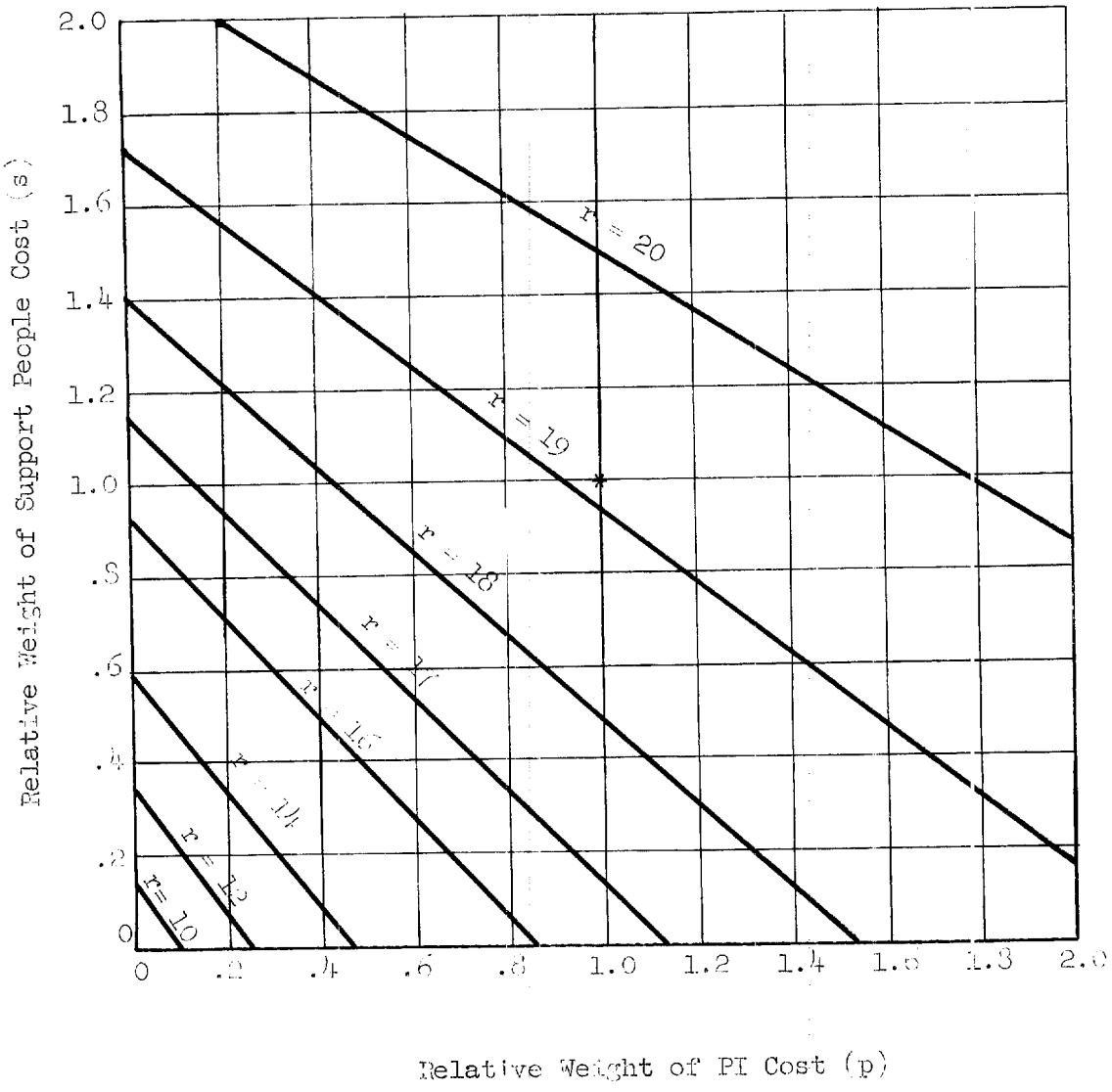


Figure 33 Effect of Varying Cost Weighting Factors

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

the weighting factors on the cost factors. This indicates that the relative cost effectiveness ratios for the two systems are relatively insensitive to changes in these weighting factors. For instance reducing both weighting factors from unity to zero only reduces the ratio of the cost effectiveness ratios from about nineteen to eight. This is because the cost factors of the two systems do not differ greatly.

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

APPENDIX A

Analysis of Routine Dissemination/Request

Given n_o sets of data and the independent probability of requiring each of P_1, P_2 --- P_{n_o}. The sets are ordered such that $p_i \geq p_i + 1$; the number of characters in each set is c_i; and the percent of targets for which the set is available is a_i. If the n most probable sets are given routinely:

1. The probability that no others will be required is:

$$P_o(n) = (1 - p_{n+1})\ (1 - p_{n+2}) \text{ --- } (1 - p_{n_o})$$

2. The probability that one or more additional sets are required is:

$$P_{1+}(n) = 1 - P_o(n)$$

3. The probability that exactly one additional set is required is:

$$P_1(n) = \sum_{i=n+1}^{n_o} p_i \frac{P_o(n)}{1 - p_i}$$

4. The probability that two or more additional sets are required is:

$$P_{2+}(n) = 1 - P_o(n) - P_1(n)$$

5. The expected number of requests for additional sets is:

$$N(n) = \sum_{i=n+1}^{n_o} p_i$$

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

The above probabilities are based on the condition that the target is read out, i.e., it is not cloud covered. If the probability that readout takes is only p_c, then they must be modified as follows to describe predicted targets.

$$p'_{1+}(n) = p_c \left[1 - P_o(n)\right]$$

$$p'_{2+}(n) = p_c \left[1 - P_o(n) - P_1(n)\right]$$

$$N'(n) = p_c \sum_{i=n+1}^{n_o} p_i$$

The time used to assimilate the data that is given routinely and that requested is assumed to be proportional to the amount of data and can be found from the following:

1. The average number of characters given routinely is:

$$C_g(n) = \sum_{i=o}^{n} a_i c_i$$

Note: The set, i = o, represents a set that might be given even if no data is required, i.e., the header.

2. The expected number of characters retrieved if the target is read out is:

$$C_r(n) = \sum_{i=n+1}^{n_o} p_i c_i$$

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

3. The total number of characters available is:

$$C_T = \sum_{i=o}^{n_o} a_i \, c_i$$

4. The time to assimilate the data given and that requested is:

$$t_a = \frac{C_g(n) + C_r(n)}{C_T} T_o$$

where T_o is the time to assimilate C_T characters.

To find the total time devoted to data, allowance must also be made for the time required to initiate requests and receive responses. This is proportional to the product of the expected number of requests and the time for each request, T_r. Therefore the total time is:

$$t = \frac{C_g(n) + C_r(n)}{C_T} T_o + T_r N(n)$$

$$\frac{t}{T_o} = \frac{C_g(n) + C_r(n)}{C_T} + \frac{T_r}{T_o} N(n)$$

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

Appendix B

Analysis of Data Base Utilization

The purpose of this analysis was to establish the total number of requests on each sector (and in some cases, parts of sectors) of each file and to sort these in order of use. Since the results will be used to assign the various types of data to the more and less accessible portions of the storage, the ordering of sectors was done on the basis of the ratio of the number of requests to sectors of a given type to the amount of storage needed for those sectors.

To simplify the manipulation of data in this analysis the data base was considered in terms of the five files below and the sectors on Table B-I.

1. COMIREX Installations (IDF)
2. NPIC Installations (IDF)
3. High Interest Objects (ODF)
4. Low Interest Objects (ODF)
5. A combination of inactive installations, Exploitation Products Data File (EPDF), and Mensuration Parameters File (MPF).

The estimated request rates were considered in terms of a three dimensional matrix where $r_{f,s,l}$ is the number of requests to file f, sector s, per unit load 1. The unit loads were taken as

1=1; 100 installation readouts - first phase
1=2; 100 installation readouts - second phase
1=3; 100 installation readouts - third phase
1=4; 100 queries from COINS

The amount of data in the data base was specified in terms of $N_{f,s}$, the number of characters in sector s of each unit record in file f, and m_f, the number of unit records in file f. One set of this input data is shown on Table B-II. Each "R VECTOR" is the set of $r_{f,s,l}$ values for indicated value of 1, i.e., 1, 2, 3, 4. The five lines in each array correspond to the five files. While the values of M are shown as a matrix, the value of M_f is the sum of the numbers in the corresponding line, i.e., line 1, 2, 3, 4, and 5.

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

TABLE B-I

DATA BASE DIVISION FOR UTILIZATION STUDIES

Files 1 & 2

Sector	
1.	Header
2.	Brief Description
3.	Detailed Description
4.	Readout Requirements
5.	Collateral
6.	PI Notes
7.	Photo References - Under 2 years
8.	Photo References - Over 2 years
9.	Source Data References
10.	Concise History - Under 2 years
11.	Concise History - Over 2 years
12.	Complete History - Under 2 years
13.	Complete History - Over 2 years
14.	Formatted Fixed Facilities
15.	Formatted Mobile Equipment Under 4 years
16.	Formatted Mobile Equipment Over 4 years

Files 3 & 4

Sector	
1.	Header
2.	Dimensional Data
3.	Object Description
4.	Photo References
5.	Graphic References
6.	Readout Requirements
7.	Source Data References
8.	Location Information

File 5

Sector	
1.	EDPF Header - Under 2 years
2.	EDPF Location Data - Under 2 years
3.	EDPF Alternate Titles - Under 2 years
4.	EDPF Header - Over 2 years
5.	EDPF Location Data - Over 2 years
6.	EDPF Alternate Titles - Over 2 years
7.	MPF - Under 3 months
8.	MPF - 4-12 months
9.	MPF - 1-2 years
10.	MPF - Over 2 years
11.	IDF - Inactive - Header
12.	IDF - Inactive - Selected Data

TOP SECRET

TOP SECRET

25X1

25X1

SCR 352

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

TABLE B-II
UTILIZATION ANALYSIS INPUT DATA

```
CLB, BUILD SAMPL

CLB, SAMPL LD XQ

ORTING PROBLEM 1

        P VECTOR
          38.  94.   6. 200.

        R1 VECTOR
         336. 153.  10. 100.  30.  55. 210.  25.  33.  10.   5.   0.   0. 219.  81.  27.
          21.   6.   1.   6.   0.   3.   5.   1.   6.   1.   0.   0.   0.   2.   0.
          66.  19.   8.  54.   3.   8.   4.  74.   0.   0.   0.   0.   0.   0.   0.
          24.   9.   4.  15.   0.   4.   0.  23.   0.   0.   0.   0.   0.   0.   0.
          10.  20.  40.   1.   1.   2.   6.   0.   0.   0.   7.   2.   0.   0.   0.

        R2 VECTOR
           5.   5.   0.   0.   0.   0.   0.   0.   0.   0.   0.   0.   0.   3.   0.
         176. 224.   1.  91.  28.  44. 271.  15.  91.   5.   3.   0.   0.  56.  28.   9.
          24.  12.   6.   4.   0.   8.   4.   8.   0.   0.   0.   0.   0.   0.   0.
          34.  10.   4.  53.   0.   4.   4.  58.   0.   0.   0.   0.   0.   0.   0.
          20.  34.  80.   1.   1.   2.   5.   0.   0.   0.  15.   2.   0.   0.   0.

        R3 VECTOR
         160. 105.  60.   0.  25.   0. 120.  20.  80.  45.  10.   4.   2.  35.  72.  24.
          35.  25.  12.   0.   7.   0.  28.   7.  20.  10.   5.   1.   1.   8.  17.   5.
          52. 133.  29.  12.  13.  20.  10.  64.   0.   0.   0.   0.   0.   0.   0.   0.
          20.  30.   6.   2.   7.   0.   6.  20.   0.   0.   0.   0.   0.   0.   0.   0.
          30.  66. 120.  15.  34.  60. 100.  20.   4.   1.   0.   0.   0.   0.   0.   0.

        R4 VECTOR
          38.  29.   6.   2.   2.   0.  20.   3.  10.  18.   2.   4.   1.  24.  28.   9.
          26.  15.   2.   3.   2.   0.  17.   4.   3.   7.   3.   1.   1.   8.  23.   8.
          33.   6.   5.   2.   1.   4.   5.  14.   0.   0.   0.   0.   0.   0.   0.   0.
          17.   4.   3.   2.   1.   2.   3.   7.   0.   0.   0.   0.   0.   0.   0.   0.
          10.  13.  20.   1.   1.   2.   5.   1.   0.   0.   0.   0.   0.   0.   0.   0.

        N MATRIX
            90.  1225.  2400.   200.   750.   200.  1570.  3400.   300.  2000.  4000.  6000. 17000.   300.  4000.  1500
            90.   750.    60.   200.   120.   100.   770.  1380.   150.   600.  1260.  2000.  3800.    50.   400.   150
           120.    30.   300.    60.    20.   200.    70.  1000.     1.     1.     1.     1.     1.     1.     1.     1
           120.    30.   300.    60.    20.   200.    70.  1000.     1.     1.     1.     1.     1.     1.     1.     1
            88.    59.   165.   728.    50.  1365.   690.  2070.  2760.  8940.    54.  1500.     1.     1.     1.     1

        M MATRIX
          5000.     1.     1.     1.
         25000.     1.     1.     1.
           500.     1.     1.     1.
          3500.     1.     1.     1.
         10000.     1.     1.     1.
```


TOP SECRET

TOP SECRET

25X1

25X1

SCR 352

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

The analysis procedure forms the appropriate products and sums these over 1 to determine the total number of requests to file f, sector s for all loads

$$R_{f,s} = \sum_{1=1}^{4} \left(r_{f,s,1}\right) \cdot \left(P_1\right)$$

The amount of storage required to satisfy these requests is

$$N_{f,s} = \left(r_{f,s}\right) \cdot \left(m_f\right)$$

The ration used to order the sectors in accordance with use is

$$\frac{R_{f,s}}{N_{f,s}} \cdot$$

A cumulative sum of the $R_{f,s}$ and $N_{f,s}$ from the beginning of the list is then formed to indicate the amount of the data base required tc answer the corresponding number of the requests. These sums are then expressed as a ratio of the maximum value of each.

The results of the analysis of the data on Table B-II is shcwn on Table B-III and the fractional cumulative totals were presented as Figure 16 (Section 5.1.2).

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

TABLE B-III
UTILIZATION ANALYSIS RESULTS

F	S	K	R(IJK)	N(IJK)	SUM R	SUM N	R RATIO	N RATIO	RAT(IJK)
3	2	1	3848.	15090.	3848.	15090.	0.0124	0.0000	0.2550
3	1	1	11676.	60360.	15524.	75450.	0.0501	0.0001	0.1934
3	4	1	2900.	30180.	18424.	105630.	0.0594	0.0001	0.0960
1	1	1	21798.	450270.	40222.	555900.	0.1298	0.0007	0.0484
3	7	1	1588.	35210.	41810.	591110.	0.1349	0.0008	0.0451
3	5	1	392.	10060.	42202.	601170.	0.1362	0.0008	0.0389
4	4	1	5964.	210180.	48166.	811350.	0.1555	0.0011	0.0283
4	2	1	2262.	105090.	50428.	916440.	0.1628	0.0012	0.0215
3	6	1	1976.	100600.	52404.	1017040.	0.1692	0.0014	0.0196
4	1	1	7628.	420360.	60032.	1437400.	0.1938	0.0020	0.0181
3	3	1	2042.	150900.	62074.	1588300.	0.2004	0.0022	0.0135
3	8	1	6748.	503000.	68822.	2091300.	0.2222	0.0029	0.0134
5	2	1	6952.	590177.	75774.	2681477.	0.2446	0.0037	0.0117
2	1	1	22752.	2250270.	98526.	4931748.	0.3181	0.0069	0.0101
1	14	1	13332.	1500900.	111858.	6432648.	0.3611	0.0090	0.0088
5	3	1	13760.	1650495.	125618.	8083143.	0.4055	0.0113	0.0083
2	14	1	6912.	1250150.	132530.	9333294.	0.4279	0.0130	0.0055
5	1	1	4440.	880264.	136970.	10213559.	0.4422	0.0143	0.0050
1	4	1	4200.	1000600.	141170.	11214158.	0.4558	0.0157	0.0041
4	7	1	1012.	245210.	142182.	11459368.	0.4590	0.0160	0.0041
4	5	1	242.	70060.	142424.	11529429.	0.4598	0.0161	0.0034
5	11	1	1676.	540162.	144100.	12069591.	0.4652	0.0169	0.0031
2	9	1	9502.	3750450.	153602.	15820040.	0.4959	0.0221	0.0025
1	9	1	3734.	1500900.	157336.	17320943.	0.5080	0.0242	0.0024
4	8	1	7846.	3503000.	165182.	20823942.	0.5333	0.0291	0.0022
1	6	1	2090.	1000600.	167272.	21824540.	0.5400	0.0305	0.0020
1	2	1	12714.	6128676.	179986.	27953212.	0.5811	0.0391	0.0020
2	4	1	9382.	5000601.	189368.	32953815.	0.6114	0.0461	0.0018
2	6	1	4250.	2500300.	193618.	35454113.	0.6251	0.0496	0.0016
1	7	1	12700.	7854711.	206318.	43308822.	0.6661	0.0606	0.0016
2	7	1	29232.	19252312.	235550.	62561126.	0.7605	0.0876	0.0015
4	6	1	928.	700600.	236478.	63261720.	0.7635	0.0886	0.0013
2	2	1	24434.	18752253.	260912.	82013981.	0.8424	0.1148	0.0013
4	3	1	1164.	1050900.	262076.	83064864.	0.8461	0.1163	0.0011
5	5	1	536.	500150.	262612.	83565011.	0.8479	0.1170	0.0010
2	5	1	3074.	3000360.	265686.	86565363.	0.8578	0.1212	0.0010
2	15	1	7410.	10001202.	273096.	96566570.	0.8817	0.1352	0.0007
2	16	1	2476.	3750450.	275572.	100317022.	0.8897	0.1405	0.0006
1	15	1	9392.	20012007.	284964.	120329014.	0.9200	0.1685	0.0004
1	5	1	1690.	3752250.	286654.	124081256.	0.9255	0.1738	0.0004
1	10	1	4250.	10006003.	290904.	134087252.	0.9392	0.1878	0.0004
2	3	1	604.	1500180.	291508.	135587436.	0.9412	0.1899	0.0004
1	16	1	2970.	7504501.	294478.	143091921.	0.9508	0.2004	0.0003
5	7	1	2298.	6902071.	296776.	149993971.	0.9582	0.2101	0.0003
1	3	1	1940.	12007202.	298716.	162001153.	0.9644	0.2269	0.0001
2	10	1	1968.	15001803.	300684.	177002963.	0.9708	0.2479	0.0001
1	8	1	1670.	17010204.	302354.	194013135.	0.9762	0.2718	0.0000
5	6	1	1024.	13654096.	303378.	207667231.	0.9795	0.2909	0.0000
2	8	1	2290.	34504147.	305668.	242171355.	0.9869	0.3392	0.0000
5	4	1	422.	7282185.	306090.	249453527.	0.9882	0.3494	0.0000
1	11	1	650.	20012007.	306740.	269465582.	0.9903	0.3775	0.0000
2	11	1	912.	31503787.	307652.	300969305.	0.9933	0.4216	0.0000
1	12	1	824.	30018007.	308476.	330987309.	0.9960	0.4636	0.0000
5	12	1	264.	15004502.	308740.	345991763.	0.9968	0.4847	0.0000
5	8	1	320.	20706215.	309060.	366697978.	0.9978	0.5137	0.0000
2	12	1	206.	50006015.	309266.	416703882.	0.9985	0.5837	0.0000
1	13	1	212.	85051020.	309478.	501754838.	0.9992	0.7029	0.0000
2	13	1	206.	95011412.	309684.	596766298.	0.9999	0.8360	0.0000
5	9	1	24.	27608285.	309708.	624374499.	0.9999	0.8747	0.0000
5	10	1	6.	89426838.	309714.	713801338.	1.0000	1.0000	0.0000

TOP SECRET

25X1

TOP SECRET

25X1

SCR 352

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1 [redacted] TOP SECRET [redacted] 25X1
Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

APPENDIX C

Tape Storage System

The purpose of this analysis is to establish the response time to requests and cost of a tape storage system. The design parameters to be considered are the fraction of the total data base (750 million characters) that is stored on on-line magnetic tapes and the particular configuration of tapes.

The response time to be considered is the storage system response time and is a properly weighted average that includes the time to get data from off-line storage and the time to respond from tape in the various failure modes as well as the time to respond in the normal mode. The average system response time is defined as

$$T_R = r \left[P_o \; T_{R_o} + \sum_i P_i \; T_{R_i} \right] + \left(1-r\right) T_m$$

where r = fraction of total requests answered from tapes
P_o = probability that tape system is operating in the normal mode
T_{R_o} = response time of the tapes in the normal mode
P_i = probability that tape system is operating in failure mode i
T_{R_i} = response time in failure mode i
T_m = response time from off-line storage

1. Tape Response Time - Normal Mode

In considering the several configurations of tape storage that might be used, first consideration was given to a "first-in, first-out" system since it generally has the shortest average response time. However, the IIS request rates (see Appendix B), of as much as 20,000 per day, cannot be reasonably handled in that manner. Therefore a batch process was considered in which the tapes are read in a fixed cycle and requests are

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

A better strategy, when there are a number of tapes in the cycle, is to have the number of times each tape is read be proportional to the square root of the fraction of the total requests answered by that tape. It can be shown that, with the IIS data base characteristic, as shown in Figure 16 (Section 5.1.2), such a strategy can produce an average response time equal to one-sixth of the time to read one tape times the number of tapes in the cycle.

The number of tapes in the cycle is determined by the total number of tapes, N_t, and the number of tape control units, N_{TC}. Therefore the response time in the normal mode can be expressed as

$$T_{R_O} = \frac{1}{6} \frac{N_t \quad T_t}{N_{TC}} = \frac{1}{6} \; n \; T_t$$

where T_t = time to read one tape (240 seconds)

Thus n, the ratio of the number of tapes to the number of tape control units becomes the design parameter that describes the tape configuration.

2. Reliability

The reliability investigation has been based on the assumption that the total computer system has, for various non-critical functions, one tape control unit and ten tape units that may be used as backup provided they are not required for more than five percent of the time. The probability of having more than ten tape units down simultaneously, even with forty tapes in the system is negligible. Therefore the reliability of tape units can be taken care of by providing enough extra tape units to allow for the average down-time of tape units, namely 5%.

The reliability consideration of the control units is somewhat different. Only one can be borrowed, and the probability of two being down simultaneously is significant. Additional backup units could be provided, but if this is done they should be used during normal operations to reduce the response time. The procedure followed was to consider the configuration as being supplied with N_{TC} control units. Under the condition of no units down, the response time is as stated above. Under the condi-

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

$$T_{R_i} = \frac{1}{6} \frac{N_t T_t}{N_{TC} - i} = \frac{n}{6} T_t \left(\frac{1}{1 - \frac{i}{N_{TC}}} \right)$$

The probability that i units out of N_{TC} are down can be found from

$$P_i = \binom{N_{TC}}{i} (a)^i (1-a)^{N_{TC}-i}$$

where 1-a is the availability of a tape control unit (98%).

The probability, P_i', of operating with i units less than N_{TC} is related to P_i by the availability of the borrowed control unit as follows:

if $P_1 \leq .05$; $P_o' = P_o + P_1$, $P_1' = P_2$, $P_2' = P_3$, etc.

if $P_1 > .05$; $P_o' = P_o + .05$, $P_1' = P_1 - .05$, $P_2' = P_2 - .05$, $P_3' = P_3 - .05$, etc.

Using these relationships, the average response time for the tape system can be expressed as

$$T_R = \sum_{i=0}^{i_{max}} \left(P_i' \; T_{Ri} \right) = \frac{1}{6} nT_t \sum_{i=0}^{i_{max}} \frac{P_i'}{1 - i/N_{TC}}$$

The summation was evaluated over the range of N_{TC} of interest and found not to differ from unity by more than 15%. Therefore, for the purposes of this analysis it will be considered unity. With this approximation, the response time of the tape system can be taken as

$$T_R = \frac{1}{6} \quad nT_t$$

3. Storage System Response Time

The fraction, r, of the total requests answered by data on the tapes is a function of the fraction, f, of the total data stored on tape. This function is given by Figure 16 (Section 5.1.2).

Thus the average system response time can be evaluated as a function of f from the following:

$$\overline{T_R} = \frac{1}{6} \quad nT_t r + T_m \left(1-r\right)$$

where T_m = time to obtain respone from off-line storage
$r = f(f)$ given in Figure 16 (Section 5.1.2)

4. Storage Hardware Cost

The hardware costs are determined by the numbers of tapes and tape control units. These can be expressed in terms of the design parameters, n and f, and other constants as follows:

$$N_t = \frac{V_0}{V_t} f$$

$$N_{TC} = \frac{N_t}{n} = \frac{V_0 f}{nV_t}$$

where V_0 = total number of characters in the data base
(750 million)

25X1 [redacted] 25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

Thus the hardware cost can be expressed as:

$$\text{Cost} = f \frac{V_O}{V_t} \left(C_t + \frac{1}{n} C_{TC}\right)$$

where C_t = cost of each tape unit ([redacted] 25X1A

C_{TC} = cost of each tape control unit ([redacted] 25X1A

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

APPENDIX D

Drum Storage System

This analysis has the same objectives as the one described in Appendix C and the same definition of average storage system response time.

1. Drum Response Time - Normal Mode

The average access time on large-volume random-access memories such as the Fastrand drum is about 0.1 seconds. Even considering two or three drum accesses per query to account for lookups in indices, this speed is more than sufficient to accommodate the IIS query rate on a "first-in, first-out" basis. (See Appendix G and Section 5.4). Therefore the response time for a query to a drum will be taken as 0.3 seconds.

2. Reliability

The problem with the drum system is to select the proper backup configuration so that the average system response time is not severely degraded by the response time in the failure modes. This problem is heightened by the feeling that the drum, at least in its present state of development has a poorer reliability than one would like to see. For this reason, an availability figure of only 92% for the drum was used in this analysis. Although no actual figures were available, it seems reasonable to assume that the drum would not be offered commercially unless the vendor were in a position to guarantee at least that availability.

In considering the backup modes for the drum system, the ten tape units and one tape control unit from the remainder of the computer system were utilized. However, since each drum is the equivalent of seven tapes in storage capacity, these available units provide backup for only one drum. Therefore, unless additional tapes are obtained, the drum system must have sufficient redundancy to provide data base operation essentially continuously, i.e., more than 99.9% of the time. The design parameter for describing the drum configuration has therefore been chosen as e, the number of drums provided in excess of the number required to hold the data assigned to drum storage.

A separate analysis was conducted to consider relying heavily on tapes for backup, but the small difference in cost between drums and tapes was easily compensated for by the greatly reduced response time with drums

25X1 [redacted] TOP SECRET [redacted] 25X1
Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

There are no generalized equations that describe the drum system as there were for the tape system. The analysis procedure is to pick a combination of f, the fraction of the total data base assigned tc drum and thus the minimum number of drums required, and e, the excess number of drums and determine performance.

The first step is to determine that the system, including the available tape control unit and seven of the ten tapes, has sufficient redundancy to provide a system availability greater than 99.9% (down time less than two hours per year of single shift operation). This can be considered in two parts, the drum control unit(s) and the drums themselves.

Since the tapes can only provide backup for one drum, and since drums cannot operate without a control unit, two or more drum control units must be provided if more than one drum load of data (125 million characters, or one-sixth of the total data base) is assigned to drum storage. Therefore two drum control units are provided. The probability that both will be down simultaneously is 0.0004 (based on an availability of 98%).

Given two drum control units as described above, the probability of having more than one plus the excess number of drums down must be less than 0.0006. This probability can be found by evaluating the summation

$$P_{e\ +\ 2\ \text{or more}} = \sum_{e+2}^{N_o + e} P_i$$

where N_o = number of drums required
e = excess number of drums
$N_o + e$ = total number of drums
P_i = probability that exactly i drums have failed

$$P_i = \binom{N_D + e}{i} (a)^i (1 - a)^{(N_D + e - i)}$$

Since $P_{i+1} < .08\ P_i$, the summation converges rapidly, and it is sufficient to test that

$$P_{e+2} = \binom{N_D + e}{e + 2} \left(a\right)^{e+2} \left(1-a\right)^{N_D-2} < 0.0005$$

Assuming that the combination of e and f, and the resulting N_D, meet the criterion above, the next step is to determine the probability that e+1 units fail, thereby requiring the use of the tapes as backup. This is found from

$$P_{e+1} = \binom{N_O + e}{e + 1} \left(a\right)^{e+1} \left(1-a\right)^{N_O - 1}$$

This is used to find the average response time for queries answered from the on-line storage as follows

$$T_R = \left(1 - P_{e+1}\right)\left(T_D\right) + \left(P_{e+1}\right)\left(3.5\ T_t\right)$$

where T_D = drum access time (0.3 seconds)

$3.5\ T_t$ = tape access time for a seven tape cycle (840 seconds)

3. Storage System Response Time

The number of drums, N_D, and the capacity of the drum, determine the amount of data assigned to on-line storage and this and Figure 16 (Section 5.1.2) determine the fraction of the total requests, r, answered by the on-line storage. The average storage system response time can be expressed as:

$$\overline{T} = \left[\left(1 - P_{e+1}\right)\left(T_D\right) + \left(P_{e+1}\right)\left(3.5\ T_t\right)\right] r + T_M \left(1-r\right)$$

4. Storage Hardware Cost

The hardware costs are determined by the numbers of drums and drum control units.

$$\text{Cost} = (N_D + e)\, C_D + N_{DC}\, C_{DC}$$

Where C_D = cost of drum [redacted] 25X1A

C_{DC} = cost of drum control unit [redacted] 25X1A

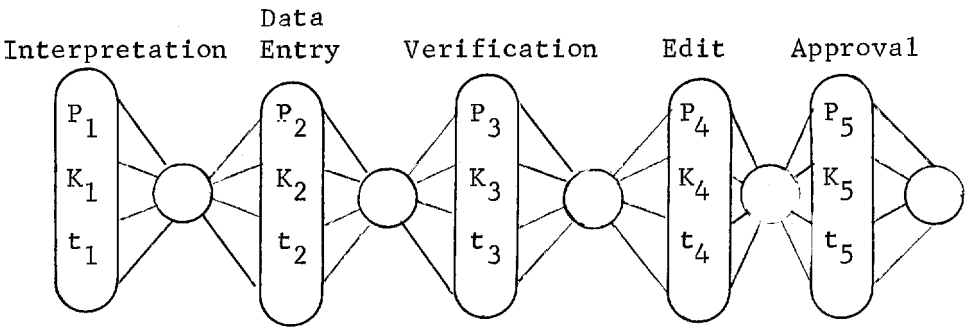
N_{DC} = number of drum control units. (This is taken as two except in those cases where $N_D + e$ exceed eight in which case three is used.)

25X1 [redacted] TOP SECRET [redacted] 25X1
Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

APPENDIX E

System Model

A simplified description of the system can be depicted by the following block diagram with each block representing one operation utilized in processing data.



The total number of people needed to perform the processsing operations can be represented by the following:

$$P_T = P_1 + P_2 + P_3 + P_4 + P_5 = \sum_{n=1}^{5} P_n$$

where P_n = number of people in the n^{th} operation and the subscripts, 1,...5, indicate the individual operations of interpretation, data entry, verification, edit, and approval, respectively.

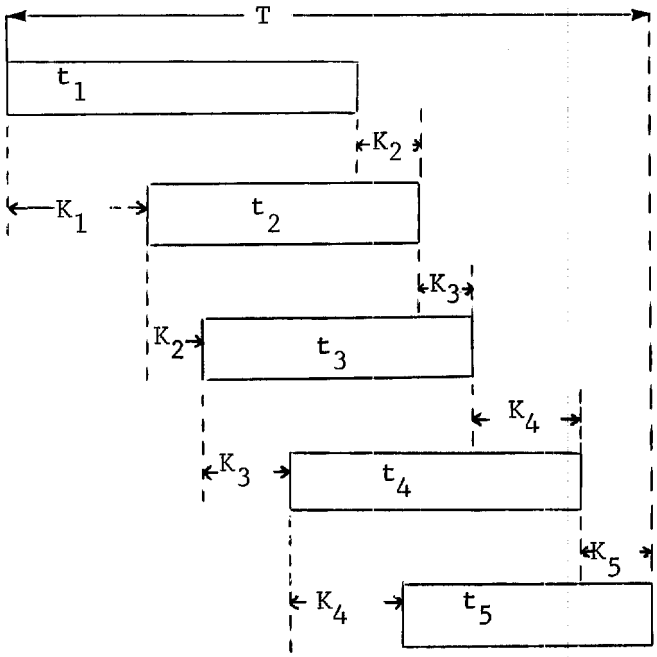
Corresponding to each operation is a t_n, total time per person utilized in the n^{th} operation. By defining N as the number of units to be serviced within the set of operations and K_n as the time to service one unit in the n^{th} operation, then

$$P_n = \frac{N K_n}{t_n}$$

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1



One useful description of total time, T, is

$$T = t_1 + K_2 + K_3 + K_4 + K_5$$

From the above definition of P_n, the total number of people

$$P_T = \frac{NK_1}{t_1} + \frac{NK_2}{t_2} + \ldots + \frac{NK_5}{t_5} = \sum_{n=1}^{5} \frac{NK_n}{t_n}$$

where $t_n = t_1 - K_1 + K_n$

The interpretation, or first operation, is performed by a group of PI's and the remaining four operations are performed by support people. Because the unit service times in support operations are small compared to that in the interpretation operation, the equation for the total number of people

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

where $K_z = \frac{K_2 + K_3 + K_4 + K_5}{4}$

and $K_z << K_1$

25X1 [redacted] TOP SECRET Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3 [redacted] 25X1

APPENDIX F

Reliability vs. Lost Time

Because the use of consoles for the support processes is being considered, their reliability or the effect of losses on the time to complete a process is investigated. The model used is the same as that used in Appendix E.

The expected number of failures per process, per year, F, can be determined by

$$F = \frac{B\ NK_n}{60\ MTBF}$$

where B = number of buckets of film per year
MTBF = mean time between failures (hr.)
N = number of units to be serviced
K_n = time to service one unit in the n^{th} process (min.)

The time lost per loss of console per process Δt_n, is defined as

$$\Delta t_n = \frac{NK_n}{P_n - 1} - \frac{NK_n}{P_n} = \frac{NK_n}{P_n(P_n - 1)} \cong \frac{NK_n}{P_n^2}$$

Another way of looking at it is

$$t_n = \frac{NK_n}{P_n}$$

$$dt_n = \frac{-NK_n}{P_n^2}\ dP_n \quad \text{or} \quad \frac{-NK_n}{P_n^2}\ \Delta P_n \cong \Delta t_n$$

For a decrease of one console, that is $\Delta P_n = -1$, $\Delta t_n \cong \frac{NK_n}{P_n^2}$

Time lost per process per year, $\bar{T} = \Delta t_n F$, or

25X1 [redacted] TOP SECRET [redacted] 25X1
Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

APPENDIX G

Queues and Delays

To investigate possible delays in retrieving information from the system caused by the variable request loads over a period of time, principles of queueing theory or delay phenomena have been applied.

A single channel system is considered with service on a first come - first served basis. The various states of the system can be characterized by the total number of units in the system, the one in service and the number in queue.

The mean waiting time, t_w, in a system in which the arrivals or requests follow Poisson's law and the service time interval has an arbitrary distribution is

$$t_w = \frac{1}{\lambda}\left[\rho + \frac{\lambda^2\sigma_s^2 + \rho^2}{2(1-\rho)}\right]$$

where λ = arrival rate

$\rho = \frac{\lambda}{\mu}$ = load factor or utilization factor

μ = service rate

σ_s^2 = variance of service time interval distribution

This mean waiting time includes the service time interval. The mean waiting time in queue, t_q, (excluding service interval) is

$$t_q = \frac{1}{\lambda}\left[\frac{\lambda^2\sigma_s^2 + \rho^2}{2(1-\rho)}\right]$$

The mean number of units, L, in a system is represented by

$$L = \lambda t_w$$

and the mean number of units, L_q, in queue is determined by

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

$$t_w = \frac{1}{\lambda}\left[\rho + \frac{\rho^2}{2(1-\rho)}\right]$$

and

$$t_q = \frac{1}{\lambda}\left[\frac{\rho^2}{2(1-\rho)}\right]$$

If the service time interval distribution is exponential, $\sigma_s^2 = \frac{1}{\mu^2}$

Then

$$t_w = \frac{\rho}{\lambda(1-\rho)}$$

and

$$t_q = \frac{\rho^2}{\lambda(1-\rho)}$$

If ρ is small, there is little reduction in mean queue size going from exponential service to constant service, and a reduction approaching 1/2 as ρ approaches 1. This is illustrated in Figure G-1.

If a reduction in delay for a particular class of units is desired, a priority of acceptance into service can be assigned. The reduction of delay for these units is appreciable if they constitute only a small fraction of the total arriving units and/or if their mean service time is smaller than the mean service time for other classes.

For k different arrival types, priorities are assigned according to class index i, where 1 is served before 2, which is served before 3, etc. The i^{th} class of Poisson arrivals has mean arrival rate λ_i, and mean service time $\tau_i = \frac{1}{\mu_i}$ with variance σ_i^2.

Mean delay in queue

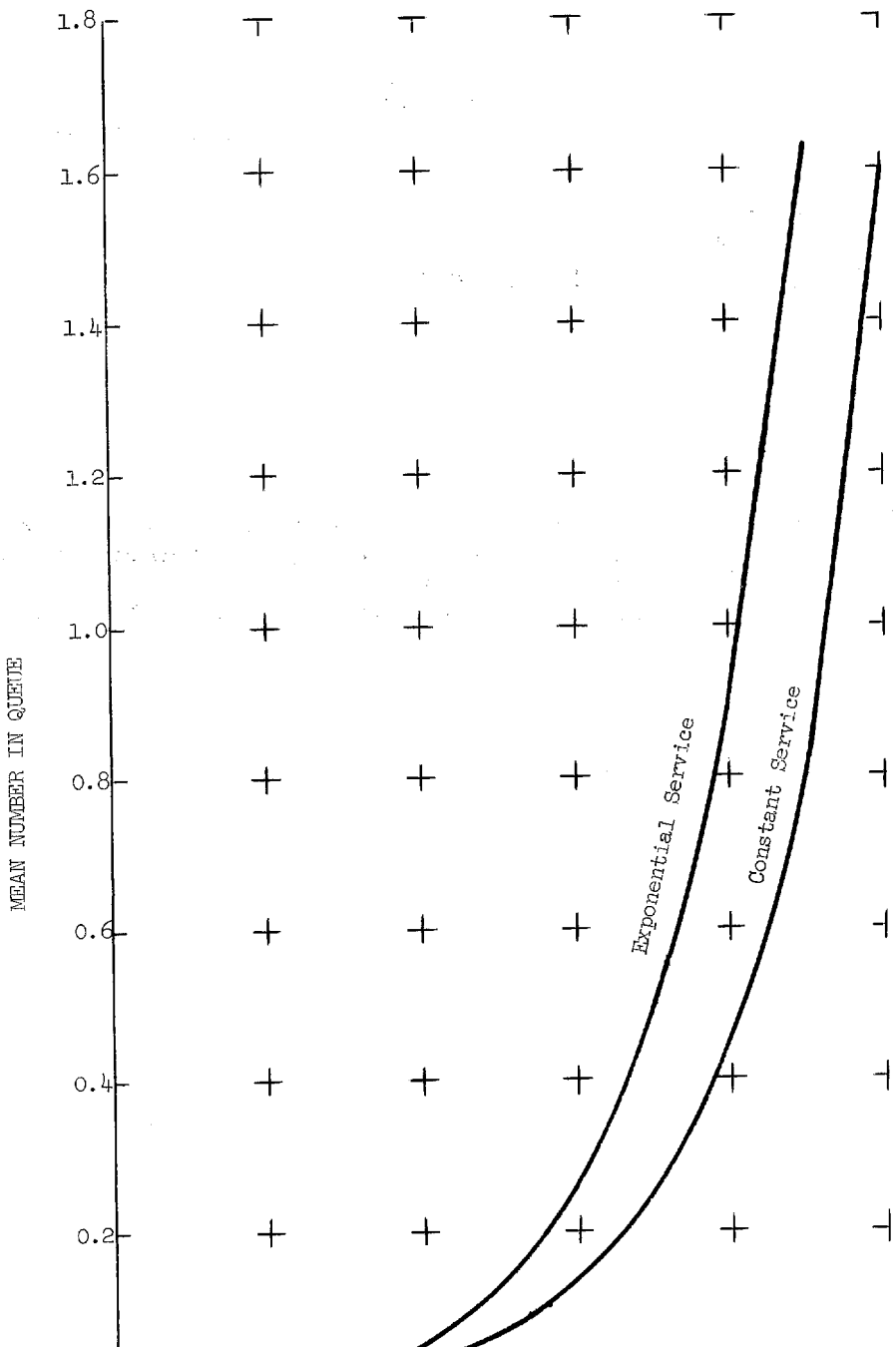
$$W_{qi} = \frac{T}{(1-\rho_{i-1})(1-\rho_i)}$$

25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

SCR 352



25X1

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

25X1

where $\rho_i = \sum_{n=1}^{i} \lambda_n \bar{\tau}_n = \sum_{n=1}^{i} \frac{\lambda_n}{\mu_n} \qquad 0 < i \leq k$

$\rho_0 = 0$

$\bar{T} = \frac{1}{2} \sum_{n=1}^{k} \lambda_n (\bar{\tau}_n^2 + \sigma_n^2)$

ρ_i, which determines the size of W_{qi} and L_{qi}, is the sum of the utilization factors $\frac{\lambda_n}{\mu_n}$ of all classes with equal or higher priority than the i^{th}. As usual with a single channel system, the mean wait in the system and the mean number in the system, for the i^{th} class, are

$$W_i = W_{qi} + \bar{\tau}_i$$

and

$$L_i = \lambda_i W_i$$

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3

Approved For Release 2004/06/29 : CIA-RDP78B04747A001900050004-3